EBV Genome Replication and Genetic Diversity in B-cells

Yi-Chun Imogen Lai

Submitted to University College London for the degree of Doctor of Philosophy

April 2012

Centre of Virology

Department of Infection and Immunity

Cruciform Building

University College London

Declaration

I, Yi-Chun Imogen Lai confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated.

Signed:.....

Date:.....

Abstract

Epstein-Barr Virus (EBV) is a human gamma herpesvirus that infects about 95% of world population. The life cycle of EBV can be divided into two stages: latent and lytic. During latency, the virus DNA is maintained as a nuclear episome with minimal protein expression to avoid immune recognition. Although primary EBV infection in childhood is usually asymptomatic, infection with EBV is associated with a wide range of proliferative diseases such as post-transplant lymphoproliferative disease (PTLD), Burkitt's lymphoma (BL) and primary effusion lymphoma (PEL). Despite EBV infection being common, less is understood about the host control of the lytic virus life cycle or the sequence diversity of the viral genome.

This thesis investigates the reactivation of lytic EBV replication in both PEL and BL cell lines and whole genome EBV sequence diversity. X-box binding protein 1 (XBP-1), a transcriptional activator that is essential for plasma cell differentiation, reactivates Kaposi sarcoma herpesvirus (KSHV), a closely related gamma herpesvirus from latency. With KSHV and EBV co-infection in PEL, the possibility of XBP-1 also reactivating EBV is investigated here. We show XBP-1 does not induce EBV into the lytic cycle in tumour B-cell backgrounds, either with or without protein kinase D. This contrasts previous observations. The lytic replication of EBV generates sequence diversity. To further understand EBV genomes and their association with the malignancies, a pipeline for sequencing EBV whole genomes using Next Generation Sequencing (NGS) has been developed. EBV whole genomes from mature B-cell tumour lines, such as PEL and BL have been sequenced assembled along with EBV from blood samples of PTLD patients. Examining the EBV whole genome sequences from PTLD blood samples through time suggests dynamic EBV sequence evolution can be observed *in vivo*. We conclude that EBV genomes contain higher variation than previously expected, and further study is needed to understand the relationship between EBV and the diseases.

Acknowledgement

Firstly, I would like to thank my supervisor Paul Kellam, for his continuous support and guidance throughout the Ph.D, and the opportunity that was provided me with. Thank you for all the encouragement throughout the study even during the down period.

I would like to thank many people, that without them, I would not be able to finish the study. For reagents I would like to thank Shanon Kenney for the gifts of Rp luciferase plasmid and BRLF1 expression vectors; Paul Farrell for the Akata, Mutu I cell lines, Zp luciferase plasmids, BZLF1 expression vectors, BZLF1 polyclonal antibody; Don Ganem for the RTA polycolonal antibody; Judy Breuer for the PTLD samples.

I would like to thank Dan Depledge, Samit Kundu, Anne Palser and Simon Watson for helping me at the analysis of the Next Generation Sequencing data and also the phylogenetic study. Thank you for sitting with me for hours and teaching me on how to carried out computation analysis, without you, the last part of this thesis will not be possible. Special thanks to Ed Tsao, Eve Coulter and Dan Frampton, you are the best group of people to work with and I really appreciate of all of the help and discussions both scientific and social. Big thanks to Shalene Singh-Shepherd and Norman Dillon for going through my thesis for me. Special thanks go to: Meleri Johns, Inga Dry, Adam Fletcher, Choon Tan, Sam Wilson, Claire Pardieu, Ben Webb, Lucy Dalton-Griffin for all the technical advice and assistance, and also everyone in the Kellam group Sanger Institute, Towers, Deenan, Breuer and Chain groups. And also, to Pamela Mansfield, Carolyn McElvaney and Lauren Collins for helping me sort out all the administrative issues. To Alice and Marcus, thank you for proofreading my thesis. To Matt, thank you for encouraging me to stay on the Ph.D program; and to my sisters Nicole and Philippa, thank you for listening to the rant and helping me through all the stressful periods.

Finally, the biggest thanks to my parents, without them, none of this would have been possible. Thank you for the support and understanding (despite continuously telling me that it would be difficult for me to find a husband with a Ph.D).

Abbreviations

AID	Activation induced cytidine deaminase
AIDS	Acquired immunodeficiency disease
APC	Antigen presenting cell
ASK-1	Apoptosis signal regulating kinase 1
ATF6	Activating transcription factor-6
BART	BamHI-A rightward transcript
BASP	B-cell specific activation protein
BCBL	Body cavity based lymphoma
BCL-6	B-cell lymphoma 6
BCR	B-cell surface receptor
BL	Burkitt's lymphoma
BLIMP-1	B-lymphocyte induced maturaton protein 1
BWA	Burrow-Wheeler Alinger
bZIP	Basic-leucine-zipper
C	Constant
CAM	Cell-adhesion molecule
CDS	coding region
CHL	Classic Hodgin's lymphoma
CHOP	C.EBP homologous protein
CRO6	Cro-Ap6
CSR	Class switching recombination
CTL	Cytotoxicity T cell
CIDR1α	Cysteine-rich interdomain region 1α
D	Diversity
ddNTP	Dideoxynucleotides
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulphoxide
DS	Dyad symmertry element
dsDNA	double stranded DNA
EA-D	Early antigen diffuse
EBER	Epstein-Barr encoded RNA
EBNA	Epstein-Barr nuclear antigen
EBV	Epstein-Barr virus

eIF-2	Eukaryotic initiation factor 2
ER	Endoplasmic reticulum
ERAD	ER associated protein degradation
ERSE	ER stress responsive element
EtBr	Ethedium bromide
FCRL4	Fc receptor like 4
FCS	Foetal calf serum
FDC	Follicular dendrictic cell
FR	Family of repeats
GADD34	Growth arrest and DNA damage inducible protein 34
GC	Germinal centre
H	Heavy
HCMV	Human cytomegalovirus
HDAC	Histone deacetylase
HHV	human herpesvirus
HIV	Human immunodeficiency virus
HL	Hodgkin's lymphoma
HLA	Human leukocyte antigen
HRS	Human Reed-Sternberg cells
HSC	Haematopoietic stem cell
HSV	herpes simplex virus
HVS	Herpesvirus simirae virus
IE	Immediate early
IFN	Interferon
Ig	Immunoglobulin
IL	Interlukin
IM	infectious mononucleosis
IR	internal repeat
ITAM	Immunoreceptor tyrosine-based activation motif
J	Joining
JNK	c-Jun amino-terminal kinase
Kb	Kilo bases
KD	Knock down
KDa	Kilo Dalton
KS	Kaposi's sarcoma

KSHV	Karposi sarcoma associated herpesvirus
L	Light
LB	Luria-Bertani
LCL	Lymphoblastoid cell line
LCV	Lymphocryptovirus
LMP	Latent membrane protein
Luc	Luciferase
LUR	Long unique region
MCD	Multicentric Castleman's disease
MCM	minichromosome maintenance complex
MCS	Majority consensus sequence
MEF2D	Myocyte enhance factor 2D
miRNA	micro RNA
MOI	multitute of infection
mRNA	messenger RNA
MTA3	Metastasis associated 1 famly member 3
MZ	Marginal zone
MZP	Marginal zone B-cell precursors
NaB	Sodium butyrate
NFDM	Non-fat dried milk powder
NGS	Next Generation Sequencing
NLPHL	Nodular lyphocyte predominant Hodgkin's lymphoma
NPC	Nasopharygeal carcinoma
NS	non-synonymous
nt	Nucleotide
ORC	Origin recognition complex
ORF	Open reading frame
OriLyt	Origin of lytic replication
OriP	Origin of plasmid replication
PAX5	Paired box protein 5
PDI	Protein disulphide isomerase
PEG	Polythylene glycol
PEL	Primary effusion lymphoma
PERK	PKR-line ER kinase
PfEMP1	Plasmodium falciparum erythrocyte membrane protein 1

PKC	Protein kinase C
PKD	Protein kinase D
PKR	Protein kinase RNA
PP1	Protein phosphatise 1
Prdm1	Positive-regulatory-domain-containing-1
PTK	Protein tyrosine kinases
PTLD	post-transplant lymphoproliferate disease/disorder
PVDF	Polyvinylidene fluoride membrane
q-PCR	quantitative PCR
QC	Quality control
RAG	Recombinase-active genes
RDA	Representational difference analysis
RDP	Recombination detection program
RDs	Repeated regions
Rp	BRLF1 promoter
RPMI	Roswell Park Memorial Institute
RTA	BRLF1
RV	Rhadinovirus
SCF	Stem-cell factor
SCS	Sanger capillary sequencing
SDF	Stromal cell-derived growth factor
SHM	Somatic hypermutation
shRNA	short hairpin RNA
SLC	Surrogate light chain
SLE	Systemic Lupus Erythematosus
SNP	Single nucleotide polumorphsm
ssDNA	Salmon sperm DNA
SYK	Spleen tyrosin kinase
TCR	T-cell receptor
TdT	Terminal deoxynucleotide transferase
TNF	Tumour necrosis factor
TPA	12-*O*-tetradecanoyl-phorbol 13-acetate
TR	Terminal repeat
TRADD	TNRF-1-associated death domain protein
TRAF	TNF receptor associated factors

TRF	Telomere repeat factor
TSA	Trichostatin A
UD	Undetectable
ML	Unique long
UNG	Uracil-N-glycosidase
UPR	Unfolded protein response
U_s	Unique short
U_L	Unique long
V	Variant
VCAM	Vascular cell adhesion molecule
VecCon	Vector control
VLA	Very late antigen
VPA	Valporic acid
VZV	Varicella-Zoster virus
WG	Whole genome
WGA	Whole genome amplification
WHO	World Health Organisation
WT	Wild type
WTSI	Wellcome Trust Sanger institute
XBP-1	X-boxed protein 1
XBP-1s	Spliced XBP-1
XBP-1u	Unspliced XBP-1
ZRE	BZLF1 response element
Zp	BZLF1 promoter
ZTA	BZLF1

Table of contents

Chapter 1

Introduction

1.1. Human Herpesviruses

The *Herpesviridae* are a large family of DNA viruses that infect nearly all animals, including humans. Over a hundred herpesviruses have been identified with hosts ranging from primates to oysters, with most species being susceptible to be infected by at least one type of herpesvirus. The name "herpesvirus" is derived from "*herpein*" in Latin meaning "to creep". This highlights one of the most important features of this family of viruses, which is that infection with herpesvirus cannot be eliminated from the hosts. The viruses have the ability to cause latent infections and as such remain in the infected host for life. The latent viruses can remain silent in the infected host, with the ability to reactivate and produce infectious virus particles. All herpesvirus are double stranded DNA (dsDNA) viruses with genomes ranging between 125-290 Kilo bases (Kb) and a characteristic virion structure (see chapter 1.1.3.1 for more detail). Currently, eight human herpesviruses (HHVs) been isolated and identified, which are divided into three sub-families; alpha-, beta- and gamma-. The sub-family and diseases associated with these eight HHVs are shown in table 1.1.1.

This Ph.D study focuses on the gamma sub-family of HHVs consisting of two members: Epstein-Barr virus (EBV/HHV-4) and Kaposi's sarcoma associated herpesvirus (KSHV/HHV-8). Gamma-herpesviruses are characterised by their limited host range, lymphocyte tropism and sequence homology. EBV and KSHV belong to the gamma-1 herpersvius (or lymphocryptovirus (LCV) and the gamma-2 herpesvirus (or rhadinovirus (RV)) respectively; the details and difference of these two viruses will be discussed later in the chapter. The gamma HHVs are the only two viruses that have been associated with tumour formation in different cellular backgrounds, and both have also been associated with more than one type of lymphoma (table 1.1.1).

Table 1.1.1. Summary of human herpesviruses (HHVs) and the associated diseases

Sub-family	Human herpesvirus	Common name	Disease associations
Alpha	HHV-1	Herpes-simplex virus (HSV)-1	Oropharangeal herpes (cold sores)
			Genital herpes
Alpha	HHV-2	Herpes-simplex virus (HSV)-2	Genital herpes
Alpha	HHV-3	Varicella-Zoster virus (VZV)	Varicella (chickenpox)
			Zoster (shingles)
Gamma	HHV-4	Epstein-Barr virus (EBV)	Infectious mononucleosis
			Nasopharyngeal carcinoma
			Burkitt's lymphoma
			Classical Hodgkin's lymphoma
			Post-transplant lymphoproliferative diseases
Beta	HHV-5	Human cytomegalovirus (HCMV)	CMV-mononucleosis
			CMV disease
Beta	HHV-6a		Exanthem subitum (sixth disease)
	HHV-6b		Roseola infantum
			Encephalitis
Beta	HHV-7		Exanthem subitum (sixth disease)
			Encephalitis
Gamma	HHV-8	Kaposi's sarcoma-associated herpesvirus (KSHV)	Kaposi's sarcoma
			Primary effusion lymphoma
			Multicentric Castleman's disease

1.1.1. Morphology of herpesvirus

Members of the herpesviruses all share similar morphology and structure, despite the taxonomical divisions within the family. Figure 1.1.1 illustrates the structure of the virion of all herpesviruses. A typical virion consists of a core with linear dsDNA, packaged within an icosahedral capsid (Furlong *et al.*, 1972; Nazerian, 1974). The capsid is surrounded by a layer of proteins called tegument proteins which are packaged by the virus envelope that forms the outer most layer of the virion. The envelope is a lipid membrane and is believed

to be derived from patches of altered cellular membrane (Rickinson & Kieff, 1996). Viral glycoproteins are embedded within the lipid envelope, forming the "spike" look. Different herpesviruses share the same virion structure but differ in the size of the virion, and the glycoproteins embedded in the envelope.

Figure 1.1.1. A) Electron microscopic picture and B) schematic diagram of the herpesvirus morphology. Linear dsDNA forms the core of the virus, packaged into an isocohedral capsid, surrounded by tegument proteins. The outer layer of herpesvirus is the envelope, which is covered by the glycoproteins that are important for viral entry. Adapted from (Young & Rickinson, 2004).

1.1.2. Life cycles

The life cycle of all HHVs can be divided into two phases; lytic and latent cycle, this allows the viruses to establish persistent lifelong infection and onward infection of new hosts. After primary infection, the virus establishes latency in different cell types (Roizman & Whitley, 2001), depending on the virus. During latency, the viral genome exists as an episome (circular DNA) in the nucleus of the host cell and minimal viral gene expression occur thus avoiding host immune surveillance. The virus enters its lytic cycle upon appropriate cellular signalling. During the lytic phase, viral replication occurs to produce new viral particles infecting new target cells. Both KSHV and EBV, as gamma herpesviruses establish latency in lymphoid cells, specifically in memory B-cells for EBV (Babcock *et al.*, 1999; Babcock *et al.*, 1998; Souza *et al.*, 2005; Thorley-Lawson, 2001). The true latent B-cell for KSHV has yet to be determined.

1.1.3. Epstein-Barr virus (EBV)

1.1.3.1. General background

EBV was discovered in 1964 by M. Anthony Epstein, Bert Achong and Yvonne Barr (Epstein *et al.*, 1964) *via* electron microscopy of cells from a Burkitt's lymphoma (BL) biopsy. BL was first described by an Irish surgeon Dennis Burkitt, as a tumour that was common to children in equatorial africa in 1958 (Burkitt, 1958). This tumour was subsequently named Burkitt's lymphoma. BL has a distinct geographical distribution (Burkitt, 1962), however, EBV infects up to 90% of population worldwide (Rickinson & Kieff, 2006), the high prevalence earned the virus the nickname "Every Body's Virus". In developing countries EBV infection usually occurs in the early years of life, often without overt symptoms. Infection with EBV later in life, such as in adolescence or young adulthood is more common in the developed countries and often associated with a lymphoproliferative disease called infectious mononucleosis (IM) (Kuppers, 2003), also known as glandular fever. Despite primary and persistent EBV infection being mostly asymptomatic; it is known that the virus is able to transform B-cells *in vitro* and is associated with the development of several B-cell lymphomas, such as Burkitt's lymphoma (BL), Hodgkin's lymphoma, primary effusion lymphoma (PEL), post-transplant lymphoproliferic diseases (PTLD). EBV is also associated with an epithelial tumour, nasopharyngeal carcinoma (NPC) (Andersson-Anvret *et al.*, 1979). EBV was the first human virus to be classified as a tumour virus (Rickinson & Kieff, 2006). Recently, it is believed that EBV infection is also associated with higher risk of some autoimmune diseases, such as systemic lupus erythematosus (SLE) (James *et al.*, 1997; Toussirot & Roudier, 2008), rheumatoid arthritis (Toussirot & Roudier, 2008) and multiple sclerosis (Ascherio & Munch, 2000). EBV is transmitted orally, and the infectious virus can be detected in oropharyngeal secretions (saliva) from the IM patients and EBV seropositive individuals whether they were immuno-suppressed or healthy (Gerber *et al.*, 1972; Strauch *et al.*, 1974; Yao *et al.*, 1985).

1.1.3.2. Epidemiology of EBV associated malignancies

Since EBV infection at childhood often occurs with no symptoms, it is difficult to determine the epidemiology of the primary virus infection. However, some EBV associated tumours are associated with distinct geographical regions. NPC occurs most frequently in Southern China, Southeast Asia, the Maghrebi Arabic regions of Northern Africa and the Arctic and less frequently in the rest of the world (Bray *et al.*, 2008; Chang *et al.*, 2009; Chang & Adami, 2006; Hsu & Glaser, 2000). BL occurs most frequently in equatorial africa. Despite BL being diagnosed in the rest of the world, only a small proportion of the cases are associated with EBV infection, whereas most of the BL in equatorial africa are associated with EBV (Brady *et al.*, 2008; Chan *et al.*, 1995; Gutierrez *et al.*, 1992; Klumb *et al.*, 2004; Mbulaiteye *et al.*, 2009; Philip, 1985; Rochford *et al.*, 2005). The correlation between the geographic and climatic distribution of endemic BL and holoendemic malaria (*Plasmodium falciparum* malaria) have been well documented in equatorial africa (Booth *et al.*, 1967; Kafuko & Burkitt, 1970). Prior to the discovery of EBV, BL was believed to be caused by pathogens transmitted by mosquitoes. Intense exposure to *Plasmodium falciparum* and EBV infection are co-factors to endemic Burkitt's lymphoma. Details of the relationship between BL and malaria will be discussed in chapter 1.3.3.

Hodgkin's lymphoma (HL) is most common in Europe and North America, to a lesser extent in South Africa, and even less common in China or Africa. However, EBV associated HL is most frequent in Africa and South Africa, but only present in moderate levels around the rest of the world (Chang *et al.*, 2009; Dinand & Arya, 2006; Ferlay, 1992; Weiss, 2000). Previous reports linked infectious mononucleosis (IM) with HL in the 1950s. The association appears to be the strongest in young adults, however it is unclear if IM (primary symptomatic infection with EBV) or primary infection (non-symptomatic) acts as risk factor of HL. Details of HL is described in chapter 1.3.5.

The reason for the three EBV associated tumours showing three different geographical patterns is not clear. There are many factors, such as human genetic variation, environmental, or other viral factors that may influence the geographical restriction. Studies have been carried out linking polymorphisms of specific EBV genes to geographical regions mostly for NPC. Little is known

about the whole virus, the association with the host genome, variation and association with other geographical factors.

1.1.3.3. EBV Virion and Genome structure

The general structure of herpesvirus virion has been discussed in chapter 1.1.1. The EBV icosahedral capsid comprises 162 capsomeres with the major capsid proteins 160, 47 and 28 Kilo Dalton (kDa) in size. This is similar to the major capsid protein of herpes simplex virus type 1 (Forman *et al.*, 1985; Foster, 1966). The most abundant capsid and tegument proteins of EBV are 350/220, and 152 kDa respectively, which differs from HSV-1 (Rickinson & Kieff, 2006). The major difference between EBV and other herpesviruses in virion structure is that EBV has predominance of a single glycoproteins gp350/220 in the outer envelope.

The EBV genome DNA in the viral particle is linear, double stranded, and 172 kilobases (kb) long and encodes 94 (type 1) and 82 (type 2) genes (Baer *et al.*, 1984; Rickinson & Kieff, 1996). During latency, the EBV genome exists as a circular episome inside the nucleus of the infected cell (Dambaugh *et al.*, 1980; Lindahl *et al.*, 1976). Figure 1.1.2 shows the simplified structure of both the episomal (figure 1.1.2.A) and linear genome (figure 1.1.2.B). EBV contains repeat regions, firstly a series of 0.5 kb terminal direct repeats (TRs) reside at both ends of the linear genome, which are joined in the episome (Dambaugh *et al.*, 1980), and internal repeat sequences (IRs), also known as the W repeats that divide the genome into short (U_S) and long (U_L) unique domains. These two unique domains are 15 kbp and 130 kbp long respectively. Unlike other herpesviruses, these two unique domains maintain their unique orientation relative to each other (Baer *et al.*, 1984). During latency, EBV replicates from the circular DNA form *via oriP*, known as the plasmid replication origin (Yates *et al.*, 1985). However, during the lytic cycle EBV is replicated *via oriLyt* (Hammerschmidt & Sugden, 1988). During lytic replication, the genome DNA is synthesized as long head-to-tail concatemers, which are cleaved and packaged into virus capsids.

There are two viral elements that are responsible for maintaining episomes: the *cis*-acting sequence, *oriP;* and a gene encoding a *trans*-acting protein, Epstein-

Barr Nuclear antigen 1 (EBNA-1) (Yates *et al.*, 1985). The *oriP* sequence is composed of two functional elements: the dyad symmetry element (DS) and the Family of Repeats (FR) (Baer *et al.*, 1984; Reisman *et al.*, 1985). The FR consists of multiple copies of a 30bp repeat unit, with each unit consisting of a 16bp palindromic sequence that constitutes an EBNA1 binding site (Ambinder *et al.*, 1990). It has been shown that, unlike other repetitive sequence that alter in copy number during viral replication, FR maintains its copy number throughout the process (Fruscalzo *et al.*, 2001), and is conserved with in strains of EBV. Hence the copy number of FR is distinct for each EBV strain (Ali *et al.*, 2009).

Figure 1.1.2. EBV genome shown as a double-stranded DNA episome (A) or a linear form (B), with the origins of plasmid replication (*oriP*) and lytic replication (*oriLyt*) indicated. The large solid blocks represent coding exons for each of the latent proteins and the arrows indicate the direction in which they are transcribed; the latent proteins include the six nuclear antigens (EBNAs 1, 2, 3A, 3B and 3C, and EBNA-LP) and the three latent membrane proteins (LMP 1, 2A, 2B). EBNA-LP is transcribed from variable numbers of repetitive exons. LMP2A and LMP2B are composed of multiple exons located on either side of the TR region, which is formed during the circularization of the linear DNA to produce the viral episome. The open arrows at the top represent the highly transcribed nonpolyadenylated RNAs EBER1 and EBER2; their transcription is a consistent feature of latent EBV infection. The outer long-arrowed line represents EBV transcription during a form of latency known as latency III, where all the EBNAs are transcribed from either the Cp or Wp promoter; the different EBNAs are encoded by individual mRNAs generated by differential splicing of the same long primary transcript. The inner shorter arrowed line represents the EBNA1 transcript originating from the Qp promoter during latency I and latency II. (**B**) Diagram showing the location of ORFs for the EBV latent proteins on a *Bam*HI restriction endonuclease map of the prototype B95.8 EBV genome. (Adapted from (Young & Murray, 2003) and (Tao *et al.*, 2006)).

1.1.3.4. EBV Typing

EBV can be sub divided into two types, namely type 1 and type 2, based on variation in the EBNA genes. The types are also referred to as type A and B in earlier publications. EBV type 1 was first sequenced from the B95.8 marmoset

cell line, which was infected with EBV from an IM patient (Baer *et al.*, 1984). The type 2 genome has been isolated and sequenced from a BL sample, AG876 (Dolan *et al.*, 2006). Previous literature suggests that the type 1 strain infects dominantly compared to the type 2 strain (Zimber *et al.*, 1986). However, a more recent study shows the possibility of EBV superinfection and co-infection with both strains of EBV (Apolloni & Sculley, 1994; Yao *et al.*, 1996). The previous methodologies of assessing EBV subtypes involve transforming B-cells with the EBV samples. Since EBV type 1 transforms B-cells more efficient than EBV type 2 (Rickinson *et al.*, 1987), this method can introduce bias towards EBV type 1 when sub-typing EBV. Nevertheless, EBV type 1 strains are more commonly found in Caucasian and Asians while type 2 strains are less frequent in Europe and the United States, but equally abundant as type 1 in Africa (Lucchesi *et al.*, 2008). EBV type 2 is not only less efficient in producing lymphoblastoid cell lines (LCLs) by immortalising B-cells; it also enters lytic cycle more readily than type 1 EBV (Brooks *et al.*, 2000; Buck *et al.*, 1999; Chen *et al.*, 1996).

The definition of type 1 and 2 EBV indicates that these two viruses differ in four nuclear proteins EBNA2, EBNA3A, EBNA3B and EBNA3C. The nucleotide difference between the open reading frames (ORFs) of these proteins for type 1 and type 2 are 55% (Lucchesi *et al.*, 2008), 10%, 12% and 19% respectively (Sample *et al.*, 1990). The percentage of differences is high compared to the non-coding region surrounding these ORFs, which is less than 5% (Sample *et al.*, 1990). However, the overall difference between the two types of EBV is considered minor (Rickinson & Kieff, 1996). The major differences at the EBNA2, 3A, 3B and 3C region of EBV type 1 and 2 are known, but it is not clear if there are other differences in other parts of the genome. Variations in genes such as BZLF1 have been observed (see chapter 1.1.2.4.), however, these differences have not been associated with a type-specific phenotype. However, small type-specific differences have been discovered before in EBER genes (Arrand *et al.*, 1989).

EBV type 1 transforms and establishes latency in B-cells better than type 2 *in vitro*, which is due to the differences between the EBNA2 and EBNA3 genes (Sample *et al.*, 1990). It has been shown a recombinant type 2 EBV with the EBNA2 coding sequence replaced by EBV type 1 EBNA2, acquires the

efficiency of transforming B-cells of the type 1 virus (Cohen *et al.*, 1989). There are some genes that are induced more strongly by type 1 EBNA2 than type 2 EBNA2, such as latent membrane protein 1 (LMP1), which is important in allowing B-cells to survive, after being driven into cell cycle by EBNA1 and –LP, however, this effect is only transient. Another gene that is induced more strongly by EBV type 1 is CXCR7, which is required for proliferation of LCLs (Lucchesi *et al.*, 2008). CXCR7 has been shown previously to be essential for transforming endothelial cells by KSHV (Raggo *et al.*, 2005).

1.1.3.5. Variation and polymorphisms of the EBV genome

Little is known about the variation of the EBV whole genome sequence. To date the variation between genomes has been specifically associated with specific genes but not the whole genome. Other than the variations associated with EBV sub-typing, polymorphisms of specific regions or genes, such as LMP1 and BZLF1, have been have been studied extensively.

LMP1 contains two commonly reported polymorphisms, one is a 30bp deletion in the C-terminus, and the other is the loss of restriction site XhoI in the N-terminus of the gene (Edwards *et al.*, 1999; Nagamine *et al.*, 2007; Saechan *et al.*, 2006; Sung *et al.*, 1998). It has been suggested that losing the 30bp region in the C-terminus results in a more aggressive phenotype of NPC disease progression and increased oncogenesis (Kingma *et al.*, 1996; Knecht *et al.*, 1993). This deletion has also been reported to be a major characterristic of EBV strains circulating in Eastern Asian (Edwards *et al.*, 1999), and Japan (Itakura *et al.*, 1996; Mori *et al.*, 1999). The majority of samples containing the loss of *XhoI* site are from Asia, and show higher frequency amongst cancer cases than controls within this region (Chang *et al.*, 2009). Despite the similarity in distribution and oncogenesis, the relationship between these two variations is unclear.

The polymorphisms between EBNA2 and EBNA3A, B and C have been used to determine the types of EBV, either type 1 or 2 (Sample *et al.*, 1990). The three EBNA3 genes share a similar gene structure. The identified conserved region of EBNA3 C terminus contains three direct repeats, which remain conserved between EBV type 1 and 2, although the number of repeats can vary in the

variants (Sample *et al.*, 1990). To date, there are five EBNA3 genes subtypes that have been determined from the EBV type 1 genome in addition to the B95.8 reference genome, with samples studied from European, region, especially the Netherlands and Austria (Gorzer *et al.*, 2006). With the five variants, EBNA3 variant 5 is the most divergent but also shares the most homology with EBV type 2 EBNA3 genes. Previous studies have shown that intertype recombination within the EBNA3 genes region (Midgley *et al.*, 2000), and the region surrounding the EBNA3A C terminus contain the active locus for EBV recombination (Gorzer *et al.*, 2006; Midgley *et al.*, 2003; Midgley *et al.*, 2000). The basic residues stretched prior to the leucine zipper domain in the N terminus of the EBNA3C allows the EBNA3C protein to maintain interaction with RBP-Jκ *in vivo* (Radkov *et al.*, 1997; West *et al.*, 2004). The variants 3 and 5 both have contain this change in this basic domain, with Arginine being replaced by Lysine, although the final charges of the region remain the same. The variant 4 contains only three out of the four Leucines in the zipper region, with Methionine replacing one Leucine (Gorzer *et al.*, 2006). The variation between EBNA3 genes has also been linked to LMP1 variants, which mainy influence the biological functions of the virus.

EBNA1 is known to contain large variations. One study in Australia sequenced the EBNA1 gene from 26 healthy individuals and 17 IM patients, resulting in 22 distinct EBNA1 sequences with 19 different protein sequences (Bell *et al.*, 2008). Two major types of variation can be based on the C-terminal region, named prototype (P) and the variant (V). Both subtypes can be further classified based on the amino acid at codon 487, and further divided into sub-variants depending on the amino acid change (Sandvej *et al.*, 2000; Wang *et al.*, 2003). Further studies have also showed variation at the N-terminus, namely at codon 16, 24 and 27 (Gutierrez *et al.*, 1997; Habeshaw *et al.*, 1999).

The variants of BZLF1 have been identified in both coding (CDS) and non-CDS regions. Variation in the non-CDS region involves differences in the number of repeats in the first intron (Packham *et al.*, 1993). Five variants have been identified from the CDS (Ji *et al.*, 2008) in comparison to the EBV type 1 reference genome. The promoter region of BZLF1 (Zp) also shows variation. Zp is a 220bp element located upstream of BZLF1 gene, containing regulatory domains to control BZLF1 transcription (see figure 1.1.5). Five Zp variants have

been identified: Zp-P represent the EBV type 1 reference sequence, Zp-V3, Zp-V4, Zp-PV (Gutierrez *et al.*, 2002; Martini *et al.*, 2007) and Zp-V1, which is the most recently detected Zp variant (Jin *et al.*, 2010). Zp-P is more common in the Americas and Europe than in China, which was believed to have predominantly Zp-V3 and Zp-V4 variants (Gutierrez *et al.*, 2002; Martini *et al.*, 2007). However, a recent study by Jin *et al* (2010) on Chinese children indicates Zp-P is the most prevalent strain in China, not Zp-V3 or Zp-V4. Zp-V3 is more common within cancer cases across geographical regions, and has been associated with malignancy in both immnocompetent and immunodeficient individuals (Gutierrez *et al.*, 2002; Martini *et al.*, 2007). More than one type of Zp variant has been found in the same individual suggesting superinfection of multiple strains of EBV.

Different EBV variants have also been found in different sample types from the same individual (Chen *et al.*, 1996; Edwards *et al.*, 2004; Gutierrez *et al.*, 1997; Sacaze *et al.*, 2001; Triantos *et al.*, 1998; Walling *et al.*, 2003), making comparison of the data more difficult. Despite many studies showing geographical association with certain polymorphisms, it is important to bear in mind the heterogeneity in the study designs, non-random sampling and the sparsity of data for many of the geographical and disease subgroups evaluated, means interpretation should be tentative.

1.1.3.6. Infection and Entry

EBV can infect a variety of cell types, such as T-cells, NK cells, smooth muscle cells (Rickinson & Kieff, 2001) and monocytes (Savard *et al.*, 2000), although B lymphocytes and epithelial cells are the two major target cells of infection. Figure 1.1.3 illustrates the process of EBV infection of B-cells. During the primary phase of EBV infection of B-cells, the B-cell specific receptor 2 (CR2) (also known as CD21), a cellular receptor for the C3d complement component, is recognized by EBV major membrane glycoprotein gp350/220 (Fingeroth *et al.*, 1988; Fingeroth *et al.*, 1984; Tanner *et al.*, 1987). This interaction is responsible for viral attachment to the cell surface (Ahearn *et al.*, 1988). gp350/220 has a dual nomenclature with genes expressed in two alternatively spliced forms, of 350 and 220 KDa each (Beisel *et al.*, 1985; Hummel *et al.*,

1984). The splice does not remove the CR2 binding domain, which is at the amino terminal part of the protein; and therefore does not affect the interaction with CR2. The interaction between CR2 and gp350/220 initiates endocytosis of EBV (Tanner *et al.*, 1987) and fusion with the plasma membrane occurs leading to virus entry. CR2 is a B-cell surface receptor; it interacts with the cellular ligand C3d inducing production of the bioactive fragment of complement protein 3 (C3), which is deposited on the surface of foreign particulate antigens during infection. CR2 then generates downstream signals to stimulate the humoral response (Guthridge *et al.*, 2001; Sarrias *et al.*, 2001; Szakonyi *et al.*, 2001; Weis *et al.*, 1984). EBV binding to CR2 blocks the interaction with C3d, stopping the signal pathway of C3 and the consequent humoral response. Therefore the gp350/220 and CR2 interaction not only enables attachment of EBV to B-cells, but also blocks the immune response.

Fusion of the EBV envelope to the B-cell plasma membrane requires other virus membrane glycoproteins including gB, gH, gL and gp42. A core fusion protein, gB, which is 857 amino acids residues in size, allows virus fusion (Haan *et al.*, 2001), but is also important for virus assembly (Lee & Longnecker, 1997). The other two glycoproteins gH and gL form a non-covalently linked complex in the virus (Li *et al.*, 1995). Removing this complex results in EBV binding but not penetrating the cells. Protein gH is the largest, with 708 amino acid residues, and gL is a 137 amino acid residue glycoprotein which is approximately 25 kDa, and remains anchored in the envelop by an uncleaved signal sequence (Li *et al.*, 1995; Yaswen *et al.*, 1993). EBV gH is dependent on gL for folding and transportation through the cell.

EBV uses different pathways to infect epithelial cells and B-cells (Hutt-Fletcher, 2007; Longnecker, 2009). To infect both cell types, gB, gH and gL are required; but to infect B-cells gp42 is also required. gp42 is a 223 amino acid protein which does not depend on the gH/gL complex for fusion with the cellular membrane. It interacts with the gH/gL complex at residue 40 and 58, with a trans membrane anchor which lies between residue 7 and 28 (Ressing *et al.*, 2005; Wang & Hutt-Fletcher, 1998). The carboxyl terminal of gp42 interacts with the variable region of the beta chain of HLA class II (Spriggs *et al.*, 1996). The interaction between gp42 and HLA class II is essential for EBV infection, which has been shown by either blocking gp42 interaction with HLA class II or by

preventing HLA class II from interacting with gp42 (Li *et al.*, 1997; Miller & Hutt-Fletcher, 1988). Binding to HLA class II by gp42 also triggers the fusion process of EBV entry of B-cells. The infectivity of EBV alters depending on the production cell types, due to the level of gp42 in the viruses produced. This results in switching of the preferred tropism of infection (Borza & Hutt-Fletcher, 2002). EBV produced by epithelial cells (E-EBV) contains higher amounts of gp42, therefore it is more efficient in infecting B-cells than epithelial cells. Moreover, the EBV produced by the HLA class II positive B-cells (B-EBV), contains less gp42 than E-EBV, and therefore is able to infect epithelial cells more efficient than B-cells. The switch of tropism of infectivity and level of gp42 in produced viruses is associated with the existence of HLA class II. This suggests that the infection of epithelial cells act as transient event, which allows a quick turnover of producing EBV that targets B-cells. B-cells releasing epithelial primed EBV into saliva also allows efficient infection of epithelial cells in a new and naïve host (Borza & Hutt-Fletcher, 2002).

EBV enters B-cells by endocytosis, internalised into cytoplasmic vesicles (Carel *et al.*, 1990; Nemerow & Cooper, 1984; Tanner *et al.*, 1987) with the virus envelope fused with vesicle membrane, releasing nucleus capsid and tegument proteins into the cytoplasm. This process takes place in a low-pH compartment (Borza *et al.*, 2004; Miller & Hutt-Fletcher, 1992). However, EBV infection of Raji cells (a BL cell line) is vesicle independent, which is believed to be due to cytoskeleton abnormalities of the tumour cells (Seigneurin *et al.*, 1977). EBV infection of epithelial cells also use a non-endocytotic mechanism (Miller & Hutt-Fletcher, 1992) at a neutral pH. Little is known about EBV dissolution and genome transport to the cell nucleus. By studying models from other DNA viruses, it is believed that the cytoskeleton network is likely to mediate EBV capsid transport to the nucleus (Dales & Chardonnet, 1973). A recent study by Valencia and Hutt-Fletcher (2012) has shown that actin is important in trafficking EBV in B-cells, and in an epithelial cells. EBV requires both actin and the microtubule network to successfully deliver EBV to the nucleus.

Figure 1.1.3. The infection and replication cycle of human herpesviruses (adapted from Mettenleiter, 2004). Schematic diagram illustrates the herpesvirus life cycle from infection to either latency or production of new virion through lytic replication.

1.1.3.7. Latency programs

EBV displays five different gene expression patterns, with one expressed during lytic reactivation and the other four expressed during the latency stage, depending on the stage of development of the infected B-cells. These are known as latency 0 (latency program), I, II (default program) and III (growth program). Different latent genes are expressed in the four latency programs. These programs were discovered in various cell lines (Amon & Farrell, 2005) and have been linked to the stage of primary EBV infection *in vivo*. Table 1.1.2 shows the details of the four latency programs and the lytic program. It is understood that the EBV genes expressed during latency control B-cell development by mimicking the B-cell signalling pathways. EBV also uses B-cell transcription factors to regulate its own gene expression.

Table 1.1.2. The latent and lytic programs of EBV. EBV expresses different genes during different latency programs, associated with different tumours and cell types during B-cell transformation (adapted from (Klein *et al.*, 2010; Thorley-Lawson, 2001; Thorley-Lawson & Gross, 2004))

Transcription program		**Promoter used for EBNA**	**Cell type**	**Associated Tumour**	**Gene expressed**
Latency III	Growth program	Cp, Wp	Newly infected B-cells	Immunoblastic Lymphomas	EBNA1, EBNA2, EBNA3A, EBNA3B, EBNA3C, EBNALP, LMP1, LMP2A, LMP2B, EBER1, EBER2, BART1
Latency II	Default program	Qp	Germinal centre B-cells	Nasopharyngeal carcinoma Hodgkin's lymphoma	EBNA1, LMP1, LMP2A, LMP2B, EBER1, EBER2, BART1
Latency I		Qp	Dividing memory cells	Burkitt's lymphoma Primary Effusion lymphoma (PEL)	EBNA1, EBER1, EBER2, BART1 (PELs also express EBER and LMP2A)
Latency 0	Latency program		Resting memory cells		EBER1, EBER2, BART1
Lytic					Lytic protein cascade

EBV in the natural context manipulates B-cell development process during infection, which is achieved by the expression of different latency programs. in newly infected B-cells, the viruses express latency type III (growth) program and drives the B-cells into formation of germinal centre B-cells, at which the latency program expressed is switched from III to II. GC B-cells further develop into dividing B-cells, and the EBV latency program switches into I and finally, when the B-cells developed into resting memory B-cells, EBV enters the final latency program, latency 0. The details of how EBV B-cell process will be discussed in chapter 1.2.5.

The latency 0 program is also known as the true latency program; and it is the program that expresses the least of the EBV latent genes, with no detectable latent mRNA or proteins. This presumably allows EBV to avoid immune surveillance and remain silent in the host. This latency program has only been identified in resting memory B-cells, and currently there has been no disease linked with this latency program.

The latency I program is commonly found in EBV positive BL. This program involves the expression of EBV encoded RNAs (EBERs) and the BamHI-A rightward transcripts (BARTs). The expression of Qp promoter induced EBV nuclear antigen-1 (EBNA1) is also observed (Figure 1.1.4). However none of the other EBNAs or the latent membrane proteins -1, 2A and -2B (LMP1, LMP2A and 2B) are expressed during this program.

The latency II (default program) can be found in NPC, and EBV positive HLs. LMP 1 and 2A/B are expressed, along with the Qp-driven EBNA1. However, the other EBNA genes are not expressed (figure 1.1.4.b). The latency II program can differ depending on the expression level of LMP2. For example, in gastric carcinoma, a low level of LMP2 is expressed; and in HL, a high level of LMP1 and 2 are expressed while both define type II latency. LMP1 and 2 expression level can also vary in the same type of malignancy (Rowe *et al.*, 2009).

Latency III, also known as the growth program, is typically found in Lymphoblastoid cell lines (LCLs) and PTLD. LCLs are generated by *in vitro* EBV infection of B-cells. In this program, EBV expresses the full spectrum of the latent genes. The EBNA 1, 2, 3A, 3B, 3C and -LP are all expressed, which are

spliced from a single poly-cistronic transcript from the Cp/Wp promoter (figure 1.1.4.).

EBNA and LMP genes are both expressed during latency. Unlike the LMP genes, which are all transcribed from individual promoters; mRNA of EBNA genes are spliced from a long transcript which is primarily produced by either the Cp or the Wp (Bodescot & Perricaudet, 1986; Bodescot *et al.*, 1987; Sample *et al.*, 1986; Speck & Strominger, 1985). During latency I and II, the only EBNA protein expressed is the episome maintaining EBNA1, which is transcribed from an alternative promoter Qp (Nonkwelo *et al.*, 1996; Schaefer *et al.*, 1995) (figure 1.1.4), with the Wp and Cp silent. During latency III, all six EBNA proteins are expressed, the transcription of the genes occurs following a promoter switch. In newly infected B-lymphocytes Wp is selectively activated post infection (Alfieri *et al.*, 1991; Woisetschlaeger *et al.*, 1990). However, despite the transcript from Wp encoding for all six EBNA proteins, only EBNA2 and EBNA-LP are expressed. EBNA2 and EBNALP proteins are the first two viral proteins expressed (Allday *et al.*, 1989; Rooney *et al.*, 1989a). The expression of these two EBNA proteins leads to activation of Cp and induces the expression of all the rest of the EBNAs (Rooney *et al.*, 1992; Sung *et al.*, 1991; Woisetschlaeger *et al.*, 1991); and also upregulate the promoters of the LMP genes (Abbot *et al.*, 1990; Fahraeus *et al.*, 1990; Wang *et al.*, 1990b; Zimber-Strobl *et al.*, 1991). The switch of promoters from Wp to Cp is driven by EBNA2 and a cellular transcription factor RBP-Jκ. EBNA2 does not directly interact with DNA, and RBP-Jκ recognised EBNA2 response element in Cp. Therefore EBNA2 and RBP-Jκ together drive the induction of Cp and leads to the expression of the other EBNA and LMP genes (Woisetschlaeger *et al.*, 1991).

Although the switch of EBNA promoters from Wp to Cp has been previously described (Alfieri *et al.*, 1991; Schlager *et al.*, 1996; Woisetschlaeger *et al.*, 1991; Woisetschlaeger *et al.*, 1990; Yoo & Speck, 2000), it is unclear of how Wp becomes repressed post Cp activation. Previous studies have suggested that the methylation state of Wp is associated with the promoter switch. Wp becomes methylated post infection with reduced activity of the promoter (Tierney *et al.*, 2000b). Other studies also indicated that methylating the Wp does not initiate the Wp silencing, but merely maintaining the silent state, also, methylating downstreams of the Wp sequence has a negative affect on Cp

transcription (Paulson & Speck, 1999). An important protein that is expressed after the promoter switching is EBNA1, which is essential for maintaining EBV latency by binding to *oriP* and initiates DNA replication of the EBV episome. EBNA1 is also a transcriptional activator. *OriP* locates just upstream of Cp and when EBNA1 binds to *oriP* it upregulates transcription from Cp (Puglielli *et al.*, 1996; Reisman & Sugden, 1986; Sugden & Warren, 1989). The expression of EBNA2 also induces transcription of LMP1 gene. EBNA1 interacts with the LMP1 promoter (LMP1p) *via* RBP-Jκ or PU.1. Unlike with Cp, where the interaction with EBNA2 is completely dependent on RBP-Jκ, the interaction between EBNA2 and LMP1p is partially dependent on RBP-Jκ (Johannsen *et al.*, 1995). EBNA 2 is able to induce LMP1 transactivation without RBP-Jκ (Fahraeus *et al.*, 1990; Fahraeus *et al.*, 1993), which suggests different mechanisms are involved in transactivation of Cp and LMP1p. Previous studies have shown that other factors such as PU.1 are involved in inducing LMP1 transactivation. PU.1, a PU-box binding protein, is a B-cell specific transcription activator. It is believed that EBNA2 transactivates LMP1p *via* interaction with PU.1 protein and a POU domain protein, instead of RBP-Jκ (Sjoblom *et al.*, 1995).

EBNA1 is expressed in all of the latency I, II and III programs; however, the transcription of EBNA1 switches from Cp in latency III, to Qp in latency II and I. It is believed that the change between latency programs is controlled by the cellular transcription factor OCT2 and EBNA1 together (Werner *et al.*, 2007). Previous studies have shown OCT2, in association with members of the Groucho (Grg/TLE) family of proteins, acts as inhibitor of Cp (Almqvist *et al.*, 2005; Malin *et al.*, 2005). Werner *et al.* (2007) had proposed that Cp was switched on by EBNA1 and switched off by OCT2, in combination with Grg/TLE; whereas Qp is switched off by EBNA1, which is otherwise on. The expression level of these two proteins allows EBV to express different latency patterns (Werner *et al.*, 2007).

It is also known that the type of EBV latency program can switch when cell lines have been cultivated for a long period of time. BL cell lines normally exhibit the same type of latency program as the original tumour *in vivo,* however, after long passages of them within tissue culture, the BL cell lines can drift and tend to display type III latency (Gregory *et al.*, 1990; Rowe *et al.*, 1987). Despite the

classification of the latency programs and the gene expression during these programs, these are not definitive (Kelly *et al.*, 2002). The latency profile does not include the expression of viral micro-RNAs in the EBV infected B-cells and epithelial cells. While B-cells are able to support all three latency programs, non-B-cells only display either latency program I or II. This is because part of the Wp/Cp promoter activation in latency III is regulated by a B-cell specific transcription factor, PAX-5, which is involved in transactivating the Wp immediately after infection of B-cells (Tierney *et al.*, 2007).

The EBV positive BLs are known to establish latency I, with occasions of latency III, however, there are at least two other types of latent gene expression profiles that are associated with BL. In one case, BL shows expression of all six EBNAs but not LMP1 or LMP2, which may be the very rare consequence of EBV integration into the cellular chromosome (Kelly *et al.*, 2006). The other type of EBV latency exhibited in BL is known as the Wp-restricted BL (see figure 1.1.4.b.), which shows the expression of all the EBNAs except EBNA2, also with no expression of LMP1 or LMP2. Wp-restricted BL is common and may be found in about 15% of the endemic BL cases (Kelly *et al.*, 2002; Kelly *et al.*, 2006; Kelly *et al.*, 2005). The EBNA2 gene is not silent but deleted from the viral episome in this type of BL cells. The Wp promoter, instead of the Qp promoter, is then used to activate the EBNA gene expression, although the exact mechanism of this is still unclear. Wp-restricted BLs also express BHRF1, a homologue of cellular Bcl-2 (Cleary *et al.*, 1986; Henderson *et al.*, 1993), a viral protein not previously associated with latency (Kelly *et al.*, 2009).

Figure 1.1.4. EBV protein expression during different latency programs. A) a schematic diagram of the EBV genome, indicating the location of the coding exon regions for EBNA1, EBNA2, EBNA3A, EBNA3B, EBNA3C and EBNALP (red arrow), BHRF1 (homologue of bcl-2) (green arrow), and the latent membrane proteins LMP1, LMP2A and LMP2B (blue arrow). The origin of latent replication (*OriP*), BamHI W repeats and terminal repeats (TR) are also shown on the diagram. (B) shows the structures of viral mRNA transcripts expressed in different latency programs. Latency 0 is not shown as there is no protein expression. The promoter is shown in black thin arrow, with coding exons coloured and non-coding exons unshaded. Latency I, expressed in most BL tumours and dividing memory B-cells, is characterised by only expressing EBNA1 from Qp. Latency II is commonly found in Nasopharyngeal carcinoma and Hodgkin's lymphoma. Latency II is also found in EBV positive germinal centre B-cells. Qp transcribed EBNA1 and the LMP family proteins are expressed during latency II. The latency III program commonly found in LCLs or newly infected B-cells is characterised by expression of all latent proteins. However with the EBNA family mRNAs transcribed as a long primary transcript and generated by alternative splicing, this is either initiated from the repeated W promoter (Wp) or the C promoter upstream. BHRF1 is also expressed from Wp initiated transcripts. LMP family proteins are transcribed from separate promoters. Wp-restricted latency, the most newly discovered latency program, only expresses the EBNA1, EBNA3A, EBNA3B, EBNA3C, EBNALP and BHRF1 proteins, with no EBNA2 and the LMPs. As the name suggests, the long primary transcripts of EBNA proteins are initiated from Wp only, and this can only be found in a subset of BL tumours with EBNA2 deletions (Rowe *et al.*, 2009).

1.1.3.8. Latent genes

Unlike other genes expressed during viral replication, the EBV genes expressed in latency have no homolog in the genomes of the other human herpesvirus. There are two families of proteins that are mainly expressed during latency: Epstein-Barr nuclear antigen (EBNA) and latent membrane proteins (LMP).

EBNA1 is a DNA binding protein (Jones *et al.*, 1989; Rawlins *et al.*, 1985), and is essential for EBV persistence in infected cells during cell division (Yates *et al.*, 1984). It controls the genome partition during cell division *via* binding to the origin of latent replication (*oriP*) (Rickinson & Kieff, 2001) and to the host EBP2 protein, which is associated with mitotic chromosomes (Kapoor *et al.*, 2005; Shire *et al.*, 1999; Wu *et al.*, 2002). EBNA1 also contains the AT-hook domain, and it is proposed that *via* this domain, EBNA1 can bind to the scaffold-associated metaphase chromosomes without EBP2 (Sears *et al.*, 2004).

EBNA2 in the most variable EBV gene between EBV type 1 and type 2. Along with EBNALP, this is one of the two first viral proteins to be expressed. *In vivo*, EBNA2 is only expressed for a short period of time immediately after the virus infects the tonsillar B-cells (Kurth *et al.*, 2003) during latency III infection. The transcription of EBNA2 mRNA was originated by Wp, and then switched to Cp in latency III program. EBNA2 does not bind to DNA directly, but in a complex with PU.1 (Johannsen *et al.*, 1995) or RBP-Jκ (Grossman *et al.*, 1994; Henkel *et al.*, 1994; Yalamanchili *et al.*, 1994), it acts on both cellular and viral promoters inducing cell proliferation (Sinclair *et al.*, 1994). EBNA2 interacts with Cp in combination with RBP-Jκ, and participates in the promoter switching in latency II, and also drives the LMP gene transactivation with PU.1. Details of promoter switching and EBNA2 interaction with various promoters were discussed in detail in chapter 1.1.3.7.

The EBNA3 family of proteins consists of three members: EBNA3A, EBNA3B and EBNA3C. The three proteins share 30% sequence homology and are transcribed as alternatively spliced transcripts (Rickinson & Kieff, 2001). EBNA3A and 3C are indispensible for viral transformation of B-cells *in vivo* (Tomkinson *et al.*, 1993) and *in vitro* (Sims, 2010). However, EBNA3B is not necessary for B-cell transformation *in vivo* (Tomkinson & Kieff, 1992) or *in vitro*. EBV with EBNA3B deletion can also induce fatal lymphoproliferative disease

(Gottschalk *et al.*, 2001). A recent study by White et al. (2012) suggests that EBNA3B acts as a tumour suppressor and restrains proliferation of EBV transformed B-cells. EBNA3B is responsible for changes in host gene expression, which leads to alteration in cell surface markers, expression and growth behaviour of infected B-cells. Knocking down EBNA3B (EBNA3BKO) in humanised mice results in formation of a more aggressive and immunoevading monomorphic diffuse large B-cell lymphoma (DLBCL) like tumours. Despite the fact that EBNA3BKO helps to prime T-cells against EBV infection, EBNA3BKO B-cells also show reduced chemoattraction to T cells. This is due to reduced secretion of chemokine CXCL10, although the mechanism which associates EBNA3BKO and reduce expression of CXCL10 in unclear. The cytotoxic T lymphocytes (CTLs) used in treating EBV-specific lympoproliferated diseases consistently express CXCR3, which is the receptor for CXCL10 (Pule *et al.*, 2008); thus decreasing the expression of CXCL10 allowing developing tumours to undergo immuno escape (White *et al.*, 2012).

The EBNA3 family of proteins are transcription regulators. EBNA3C can induce both cellular and viral gene expression, such as CD21 and LMP1 (Allday & Farrell, 1994). EBNA3C does not bind to DNA directly, but targets promoters *via* association with the RBP-Jκ transcription factor (Robertson *et al.*, 1995b), which is also a DNA-binding partner with EBNA2. Previous studies have shown that EBNA3A and EBNA3C repress RBP-Jκ-EBNA2 activated transcription (Le Roux *et al.*, 1994; Waltzer *et al.*, 1996). It can also repress Cp promoter function (Radkov *et al.*, 1997) despite the transcript being induced by Cp promoter activity, thereby forming a negative feedback loop; therefore EBNA3C acts as negative regulator of EBNA2. However, recent studies using a conditional virus have shown that inactivating EBNA3C does not result in increasing the expression of both viral and cellular genes, such as LMP1 and c-myc respectively, that are regulated by EBNA2 (Maruo *et al.*, 2006) and therefore does not antagonise the function of EBNA2. The difference between these studies could be due to the fact that the earlier studies have carried out experiments in either transient or constant transfection assays. EBNA3C was not expressed at the physiological levels but overexpressed. The promoter assays were also carried in cellular background that is not B-cell. These differences in the protein expression levels and/or cell type are possible

reasons for the differences between the studies. EBNA3C can also interact with pRb to promote transformation (Parker *et al.*, 1996). EBNA3C interacts with human histone deacetylase (HDAC) 1 (Radkov *et al.*, 1999) and HDAC 2 (Knight *et al.*, 2003) indicating that EBNA3C associates with chromatin re-modelling complexes. It has also been shown that EBNA3C is involved in cell cycle regulation, by decreasing the level of p27 (Knight *et al.*, 2005a; b) and pRb (Knight *et al.*, 2005a) in rat cells. However, the pRb level does not alter in LCLs (Maruo *et al.*, 2006) in response to the expression of EBNA3C.

EBNALP, also known as the EBNA leader protein, is encoded by the leader of the EBNA mRNA, the size of which varies depending on the number of BamHI W repeats that is encoded in the virus genome (Rickinson & Kieff, 2001). Previous studies have shown that EBNALP, although is not required for transforming B-cells *in vitro*, it is required for cell proliferation (Allan *et al.*, 1992). EBNALP also enhances the EBNA2 mediated transcriptional activation (Harada & Kieff, 1997; Nitsche *et al.*, 1997; Peng *et al.*, 2004).

Latent membrane protein (LMP) 1 is only expressed in either latency II or III, and is the major transforming protein during EBV infection of B-cells. LMP1 is essential for EBV induced B-cell transformation (Kaye *et al.*, 1993; Wang *et al.*, 1985). Expressing LMP1 induces the production of cell surface adhesion molecules and activation markers (Wang *et al.*, 1990a). LMP1 expression also upregulates Bcl-2, an anti-apoptosis protein, (Henderson *et al.*, 1991) and stimulates the production of the cytokines IL-6 and IL-8 (Eliopoulos *et al.*, 1999). By functioning as a constitutively active tumour necrosis factor (TNF) receptor, LMP1 aggregates in the membrane, and interacts with the TNF receptor associated factors (TRAFs) and TNFR-1-associated death domain protein (TRADD). This activates both NF-κB and c-Jun amino-terminal kinase (JNK) (Eliopoulos *et al.*, 1999; Kieser *et al.*, 1997; Kilger *et al.*, 1998). Both LMP1 and LMP2A are important for the survival of latency III tumours, knocking down either of the latent membrane proteins leads to NF-κB down regulation and apoptosis (Guasparri *et al.*, 2008). Many of the functions of LMP1 overlap with CD40, a member of the TNFR family. CD40 is a co-receptor of activated Ig receptors, which are important for B-cell development and differentiation. By fusing LMP1 with the CD40 C-terminus, LMP1 mimics CD40 and activates NF-κB signalled growth signals (Baxendale *et al.*, 2005; Busch & Bishop, 1999).

The LMP2 gene encodes for two proteins, -2A and -2B. These two proteins only differ in the first exon, resulting in LMP2A having a 119 amino acid cytoplasmic amino terminal domain. Neither LMP2A or 2B are essential for B-cell transformation *in vitro* (Longnecker, 2000). The main function of LMP2B is to regulate LMP2A in maintaining latency (Rechsteiner *et al.*, 2008; Rovedo & Longnecker, 2007). LMP2A encodes an immunoreceptor tyrosine-based activation motif (ITAM) (Fruehling & Longnecker, 1997). When ITAM is phosphorylated, it recruits an activated src family of protein tyrosine kinases (PTKs) and Syk PTKs and with the B-cell surface receptor (BCR) mediates B-cell differentiation and proliferation. LMP2A mimics BCR signalling pathway, the expression of LMP2A in B-cell results in proliferation and survival without BCR signalling (Caldwell *et al.*, 1998).

1.1.3.9. Lytic cycle and reactivation

EBV reactivation has been extensively studied *in vitro*. It is well known that treating latently infected B-cells in tissue culture with chemicals that alter certain intracellular pathways can trigger lytic reactivation of EBV. Such chemicals include phorbol esters, such as 12-*O*-tetradecanoyl-phorbol 13-acetate (TPA); calcium ionophores; histone deacetylase (HDAC) inhibitors, such as sodium butyrate and valporic acid; and DNA demethylating agents, such as 5-azacytidine. Cross-linking the cell-surface immunoglobulin (Takada & Ono, 1989; zur Hausen *et al.*, 1979) of latently infected B-cells also induces the EBV lytic cycle.

A previous study by Laichalk and Thorley-Lawson (2005) showed that the differentiation of peripheral memory B-cells into plasma cells in the tonsils is associated with production of infectious virus particles; which can then transfer into the lympho-epithelium and secrete virus into saliva (Hadinoto *et al.*, 2009; Laichalk & Thorley-Lawson, 2005). It is suggested that the EBV positive memory B-cell returns to the Waldeyer's ring at the tonsils and adenolds, undergoes plasma cell differentiation and therefore viral replication. The viruses produced from plasma cells infect the basolateral surface of epithelial cells of the Waldeyer's ring, which continuously shed viruses into saliva.

Two viral proteins, BZLF1 (ZTA) and BRLF1 (RTA), mediate reactivation from latency *via* direct binding of EBV DNA (Amon *et al.*, 2004; Chang & Liu, 2000; Yuan *et al.*, 2006). BZLF1 is an immediate early protein and a transcription factor that initiates the EBV lytic cycle. BZLF1 belongs to the basic-leucine-zipper (bZIP) family of transcription factors, this family also contains cellular transactivators such as c-Fos and c-Jun (Farrell *et al.*, 1989; Flemington & Speck, 1990b; Kouzarides *et al.*, 1991), which can dimerise to form the AP-1 transcription factor that upregulates genes involved in proliferation and differentiation. BZLF1 contains three functional domains: the amino terminal non-acidic transactivation domain, the basic DNA binding domain and the dimerisation domain (Farrell *et al.*, 1989; Lieberman & Berk, 1990). B-cell surface receptor (BCR) cross-linking activates the BZLF1 promoter (Zp) and this leads to BZLF1 expression. The BRLF1 promoter (Rp) is then subsequently activated (Amon *et al.*, 2004). BRLF1 is also a sequence specific DNA binding transcriptional activator, with three functional domains: an acidic transactivation domain, a DNA-binding domain and a homodimerization domain (Manet *et al.*, 1991). BRLF1 activates transcription in part through direct binding to Rta-responsive elements (RREs) (Manet *et al.*, 1991). BZLF1 is the key immediately early protein in EBV (Bryant & Farrell, 2002; Countryman & Miller, 1985) and BZLF1 expression alone is sufficient to initiate the entire EBV lytic cycle in permissive cell types (Countryman & Miller, 1985; Rooney *et al.*, 1989b). Once BZLF1 is expressed it transactivates the Rp promoter leading to the expression of BRLF1, as well as transactivating its own promoter, Zp. Together these two proteins orchestrate the rest of the EBV lytic gene expression cascade (Takada *et al.*, 1986). Some studies have shown that BRLF1 can also trigger the lytic cycle cascade alone, but this only occurred in some cell lines (Ragoczy *et al.*, 1998; Zalani *et al.*, 1996). Knocking down of either BZLF1 or BRLF1 prevents the virus lytic cycle (Feederle *et al.*, 2000).

EBV DNA replication is controlled by *oriLyt*, which is different from the control of EBV episome replication during latency, controlled by *oriP* (Hammerschmidt & Sugden, 1988). BZLF1 and five other EBV proteins are essential for the function of *oriLyt* (Fixman *et al.*, 1992). *OriLyt* contains seven BZLF1 binding sites in four clusters which are essential for replication (Fixman *et al.*, 1992; Hammerschmidt & Sugden, 1988; Sarisky *et al.*, 1996; Schepers *et al.*, 1993a; 1996; Schepers *et al.*, 1993b). BZLF1 is not only the viral lytic origin binding

protein, it also facilitates the recruitment of the EBV helicase-primase complex, a lytic replication complex (Gao *et al.*, 1998).

1.1.3.10. Regulation of lytic reactivation

The BZLF1 gene consists of three exons and one intron that contains a set of 29bp repeats. The expression is derived from three exon transcripts starting from the BZLF1 promoter (Zp) (figure 1.1.5.A). The BRLF1 reading frame is upstream of Zp and both are transcribed in the same orientation. Unlike Zp, which only produces one type of transcript, Rp can lead to three different alternatively spliced mRNAs (Manet *et al.*, 1989). Two larger mRNAs encode BRLF1 and the third one produces a BZLF1/BRLF1 fusion protein called RAZ. RAZ contains the BRLF1 amino-terminus, and the BZLF1 DNA binding and dimerisation domain (Furnari *et al.*, 1994; Manet *et al.*, 1989), this fusion protein inhibits BZLF1 transactivation, although not *via* acting as the dominant-negative protein to BZLF1. During latency, the expression of BZLF1 and BRLF1 are not detectable.

The Zp can be sub-divided into various elements (figure 1.1.5.B). The ZV element, which is near the transcription start site, binds to the transcriptional inhibitor ZEB1. ZEB1 induces repressed chromatin through recruiting a CoREST-CtBP corepressor complex (Kraus *et al.*, 2001). The ZEB binding to ZV maintains Zp silencing during latency (Binne *et al.*, 2002; Feng *et al.*, 2007; Kraus *et al.*, 2001; Kraus *et al.*, 2003; Yu *et al.*, 2007b). The myocyte-specific enhancer factor 2D (MEF2D) also recruits HDAC to Zp, which maintains the promoter as an inactive chromatin structure further maintaining the silent state of Zp (Gruffat *et al.*, 2002; Liu *et al.*, 1997b). This explains how HDAC inhibitors such as sodium butyrate, can activate EBV into the lytic cycle. Rp contains CpG dinucleotides, which are highly methylated during latency (Bhende *et al.*, 2004), contributing to the silencing of Rp during latency.

There are sequence related elements other than ZV assigned to Zp, namely ZIA, ZIB, ZIC, ZID, ZII, ZIIIA and ZIIIB (figure 1.1.5.B). Many cellular factors have been identified which bind to these Zp elements. MEF2D has been shown to bind to ZIA, ZIB and ZID elements. BCR cross-linking, shown to induce EBV reactivation, leads to modification of MEF2D *via* the phosphatidylinositol 3-

kinase/Ca^{2+} pathway, allowing other factors bind to ZI elements (Bryant & Farrell, 2002). SP1 and SP3 are both ubiquitous transcription factors and can bind to ZIA, ZIC and ZID elements (Liu *et al.*, 1997a; Liu *et al.*, 1997b). CREB, ATF-1, and a ATF2/c-Jun hetrodimer can bind to the ZII element (Adamson *et al.*, 2000; Flemington & Speck, 1990c; Liu *et al.*, 1998; Wang *et al.*, 1997). The ZIII elements interact with BZLF1 (Flemington & Speck, 1990a) with ZIIIB having higher affinity binding to BZLF1 than the ZIIIA element. Binne *et al*., (2002) indicated that the ZIIIA element still has a key role in Zp activation, following BCR cross-linking activation, however this effect is mediated *via* unknown cellular factor, and not directed by BZLF1 binding (Binne *et al.*, 2002). Therefore ZIIIA is essential for Zp activation without BZLF1. ZIIIB, with the higher affinity binding of BZLF1, is not required at the very immediate stage of promoter activation after BCR cross-linking, but at the later induction stage, which happens at about the same time as BZLF1 expression. There are many factors that have been studied and previously reported to activate Zp, such as TGF-β (Fahmi *et al.*, 2000; Liang *et al.*, 2002) and XBP-1 (Bhende *et al.*, 2007; Lai *et al.*, 2011; Sun & Thorley-Lawson, 2007), which will be discussed in further detail in chapter 3.

EBV genomes are not methylated when entering the cells, but gradually became methylated during immortalisation and during establishment of latency (Li & Minarovits, 2003; Tao & Robertson, 2003). Methylation of DNA is associated with inhibition of gene expression, by association of specific methy-CpG-binding proteins with methylated DNA, and leads to transcriptional silencing and chromotin remodelling (Klose & Bird, 2006). DNA methylation is therefore involved in regulation of EBV reactivation. Previous studies suggest that BZLF1 is able to bind to DNA even is CpG methylated (Bhende *et al.*, 2004; Dickerson *et al.*, 2009; Karlsson *et al.*, 2008a), which indicates BZLF1 is able to transactivate Zp during latency when the genome is methylated. BZLF1 is able to interact directly to sequence specific DNA which is referred to as the Z response elements (ZREs) (Sinclair & Farrell, 1992), which can be sub-divided into three classes (Karlsson *et al.*, 2008b). Class I ZRE does not contain a CpG motif and therefore the interaction with BZLF1 is not associated with methylation status. Class II ZREs contain a CpG motif and is able to interact with BZLF1 when non-methylated; however the interaction is enhanced when

the ZREs are methylated. Class III ZREs contains one or more CpG motifs. This class of ZREs are not able to or poorly interact with BZLF1 when non-methylated, but are recognised by BZLF1 when methylated (Karlsson *et al.*, 2008b). Therefore, BZLF1 is able to bind to Class III ZREs when the viral genome is methylated. The ZIIIA and B domains of the Zp have been identified to interact with BZLF1 (Flemington & Speck, 1990a); however, it is unclear if these two regions contain any class III ZREs.

Figure 1.1.5. Schematic diagrams of A) BZLF1 genes. BZLF1 contains three exons and one intron, which consists numbers of 29 bp repeats. The nucleotide coordination is according to EBV type 1 reference sequence from Genbank. B) Diagram of Zp, the promoter of the Epstein-Barr virus BZLF1 gene. Cis-active regulatory sites and proteins that bind these sites are identified. The location of the Zp polymorphisms described in section X is indicated below the diagram of Zp (Miller & El-Guindy, 2002)

1.1.3.11. EBV replication in latency

Replication of EBV during latency differs from EBV replication during lytic cycle. During latency, EBV DNA is in the closed circular episome form and the genes

required for replication are silent. The replication of the EBV episome is associated with the host DNA replication. The EBV episome is packaged with the cellular histones (Dyson & Farrell, 1985), and replicates only once with the host DNA during the S phase of the cell cycle (Adams, 1987). The two copies of EBV DNA are then divided into the daughter cells during the mitosis phase (Kirchmaier & Sugden, 1995). EBV replication during latency initiates from *OriP*, and only one protein EBNA1 is involved in the process. Since EBNA1 has no helicase activity (Frappier & O'Donnell, 1991), genome replication is dependent on cellular factors. *OriP* contains two clusters of EBNA1 binding sites, in the FR and DS regions. It is believed that the 20 EBNA1 binding sites in FR are associated with tethering the viral episome to human chromosomes, allowing the replicated DNA to be retained in the daughter cells (Kanda *et al.*, 2001; Kirchmaier & Sugden, 1995). The DS region contains two pairs of EBNA1 binding sites flanked by telomere repeat factor (TRF) binding sites (Bashaw & Yates, 2001; Deng *et al.*, 2002; Rawlins *et al.*, 1985; Yates *et al.*, 2000). The TRF binding sites contribute to DNA replication efficiency and episome maintenance, and have also been shown to recruit the origin recognition complex (ORC) in an RNA dependent manner (Atanasiu *et al.*, 2006; Norseen *et al.*, 2008). The TRFs are also responsible for recruiting cellular factors involved in DNA recombination and repair (Palm & de Lange, 2008). ORC is a cellular initiation factor, which has been shown to interact with *OriP* by binding either to or close to the region of DS elements (Schepers *et al.*, 2001). Minichromosome maintenance complex (MCM) is another cellular initiation factor, that has been shown to bind to *OriP* (Chaudhuri *et al.*, 2001). ORC and MCM form a pre-replication complex, and initiate replication with *OriP*.

1.1.4. Kaposi Sarcoma associated herpesvirus (KSHV)

KSHV is a gamma-2-herpesvirus and first member of the genus *Rhadinovirus* found to infect humans (Moore *et al.*, 1996; Neipel *et al.*, 1998). The most closely related human virus to KSHV is EBV.

KSHV is the most recently discovered member of the Herpesvirus family; first identified in 1994 in AIDS associated Kaposi's sarcoma (KS) (Chang *et al.*, 1994). KS was first described at 1872 by Moritz Kaposi, a Hungarian

dermatologist, who found that six patients with multifocal brown–red or blue–red nodules or plaques on the feet and hands (Kaposi, 1872). It was initially called “Idiopathisches multiples Pigmentsarcoma der Haut” (multiple idiopathic sarcoma of the skin). However it was not until 1994, Chang *et al*., (1994) using the technique called representational difference analysis (RDA), successfully identified KSHV from AIDS associated KS samples. KSHV has been shown to be the aetiological agent of KS, the most common malignancy (10%) found in regions of sub-Saharan Africa (Dedicoat & Newton, 2003). KSHV is also associated with two types of B-cell lymphomas: Primary Effusion Lymphoma (PEL) (Cesarman *et al.*, 1995a) and multicentric Castleman’ disease (MCD) (Soulier *et al.*, 1995). KSHV is also associated with 4% of all AIDS related non-Hodgkin lymphomas (Carbone & Gloghini, 2008), thus KSHV is the most common cause of cancer among those with untreated HIV infection (Beral *et al.*, 1990; Boshoff & Weiss, 2002).

Figure 1.1.6. Schematic diagram showing the KHV/HHV-8 genome. Human cells latently infected with KSHV harbour multiple copies of the episomes. As indicated above, the circular episome represents a fusion of the terminal repeats at the both ends of the linear genome. The episome is approximately 140 kilobase pairs in length and contains open reading frames which code for viral proteins which mediate latent infection as well as modulate cellular processes (adapted from (Cotter & Robertson, 2002).

1.1.4.1. KSHV Genome

As with other HHVs, KSHV is a double stranded DNA (dsDNA) virus, with a whole genome of 165-170 Kbp (Renne *et al.*, 1996a), consisting of a long unique region (LUR) which is approximately 140-145 Kbp in length. The genome exists in circular dsDNA episome form during latency. During lytic replication, linear dsDNA genome is produced in reactivated cells and virions (Renne *et al.*, 1996b).

The LUR encodes all the viral open-reading frames (vORFs) (Russo *et al.*, 1996). The LUR is flanked at each end by a series of non-coding GC-rich terminal repeats (TRs), which are 801 base pairs (bp) long. Typically, the TR DNA content is about 20-25 Kb in total. However, the number of TRs in different genomes ranges widely between 16-75, which accounts for the variation that can be seen in the isolates of KSHV (Lagunoff & Ganem, 1997). There are 81 ORFs which are larger than 100 amino acids in LUR. 66 of these ORFs have homologues in herpesvirus simirae (HVS), another gamma-2 herpesvirus that infects old world monkeys. KSHV encodes unique proteins such as K1-K15 and also encodes 11 microRNAs (miRNAs) (Cai *et al.*, 2005; Pfeffer *et al.*, 2005; Samols *et al.*, 2005).

Unlike EBV, only two gene expression profiles have been defined for KSHV: lytic and latent. The genes expressed during KSHV latency are dependent on the cell line. For example, eight different latent genes are expressed during latency in PELs: (vFLIP/ORF71 (Low *et al.*, 2001), v-cyclin/ORF72 (Platt *et al.*, 2000), LANA/ORF73 (Rainbow *et al.*, 1997), vIL-6/ORFK2 (Staskus *et al.*, 1999), LNAN2/ORFK10.5 (Rivas *et al.*, 2001), vIRF2/ORFK11.1 (Burysek & Pitha, 2001), Kaposin/ORFK12 (Muralidhar *et al.*, 2000) and K15/ORFK15 (Glenn *et al.*, 1999); as compared to four in KS (vFLIP/ORF71 (Sturzl *et al.*, 1999), v-cyclin/ORF72 (Reed *et al.*, 1998), LANA/ORF73 (Rainbow *et al.*, 1997) and Kaposin/ORFK12 (Zhong *et al.*, 1996). Currently, there is no further sub-typing of the latency depending on the level of latent gene expression.

KSHV reactivation can also be induced by various reagents, such as TPA by activating the protein kinase C (PKC) pathway (Renne *et al.*, 1996b); HDAC inhibitors i.e. valporic acid (VPA) and sodium butyrate (NaBut) (Miller *et al.*, 1997); or by DNA methylation inhibitors (Chen *et al.*, 2001). RTA of KSHV (K-RTA) is analogous to BZLF1 of EBV, but KSHV does not encode a protein that is equivalent of BRLF1. The expression of K-RTA induces and leads to the completion of the KSHV lytic cycle. K-RTA is often referred to as the “switch” for the KSHV lytic cycle, since its expression is suppressed during latency and only activated to induce the lytic cycle. RTA is therefore necessary (Lukac *et al.*, 1998) and sufficient (Gradoville *et al.*, 2000) to activate the KSHV lytic cycle. It is encoded by the ORF50 gene and like EBV protein BZLF1, can transactivate its own promoter (Deng *et al.*, 2000).

1.1.4.2. KSHV diversity

The majority of the KSHV genome is highly conserved, but there are some regions of KSHV such as ORF K1 and K15 that are highly variable. This has lead to classification of the virus into five different sub-types A, B, C, D and E (Hayward & Zong, 2007). The general viral nucleotide differences between these sub-types are about 3%, compared to 15-30% of variation at the K1 region. The variation can reach up to 60% at the two hypervariable regions V1 and V2, at ORF K1 region (Zong *et al.*, 2002; Zong *et al.*, 1999). It is unclear what causes the variability in the K1 region,; however, there are epitopes for cytotoxic T cells (CTLs) present at the V1 region, indicating K1 variation could be due to CTL evasion (Stebbing *et al.*, 2003).

The sub-types of KSHV can be further divided depending on the ORF K15 region, with variants known as P (prototype), M (minority) (Poole *et al.*, 1999), N and Q (Zong *et al.*, 2002). However, these four variations can only be found in the KSHV sub-type A and C, and are believed to arise from recombination with a progenitor herpesvirus (Zong *et al.*, 1999). KSHV genome typing is correlated with the geographical and/or ethnic origin with the possibility that the sub-types arise due to geographical isolation of the host population (Hayward, 1999; Zong *et al.*, 2002).

1.1.4.3. Epidemiology

The prevalence of KSHV can be divided by geographical regions and also associated with HIV prevalence. North America, North Europe and Asia have the lowest prevalence of KSHV (between 1-7%); Mediterranean, Middle East and Caribbean have about 5-25% prevalence. The highest percentage (over 50%) of KSHV infection can be found in sub-Sahara Africa and the Amazon basin. The sub-types of KSHV clades A and C are predominantly isolated from individuals from Europe, the United States, Asia and the Middle East; sub-type B can be found in sub-Saharan Africa; sub-types D and E are predominately isolated from South Asia, Australia, the Pacific and Brazilian Amerindians (Biggar *et al.*, 2000; Cook *et al.*, 1999; McGeoch & Davison, 1999; Meng *et al.*,

1999; Zong *et al.*, 1999). The transmission route and risk factors for KSHV are not fully understood.

1.1.5. Interactions between KSHV and EBV

Primary effusion lymphoma is defined as a KSHV positive body cavity based lymphomas (Nador *et al.*, 1996); however, 80% of PELs are also co-infected with EBV. EBV and KSHV in dually infected PELs have been shown to interact with each other as well as with the host cell (Fan *et al.*, 2005; Groves *et al.*, 2001; Krithivas *et al.*, 2000; Xu *et al.*, 2007).

EBV is known to enhance the tumourigenecity of the dually infected PELs in SCID mice, increasing the kinetics of tumour development and the vascularised tumour growth, in comparison to the EBV negative PELs (Trivedi *et al.*, 2004). Also, dually and singly infected PELs express different sets of cellular genes (Fan *et al.*, 2005). The KSHV LANA gene activates EBV LMP1 by directly transactivating LMP1 promoter (Groves *et al.*, 2001) but reduces EBNA1 expression, and repress EBNA2 activation (Krithivas *et al.*, 2000). Previous studies showed that in dually infected PEL cell line BC1, adding both TPA and butyrate, inducing agents of both herpesviruses, only induces KSHV but not EBV into lytic reactivation (Miller *et al.*, 1997; Sun *et al.*, 1998). Under the dually infected environment, EBV BZLF1 and KSHV RTA (K-RTA) co-localise and interact *in vivo*. This interaction is associated with the leucine heptapeptide repeat region of K-RTA and the leucine zipper region of BZLF1. K-RTA inhibits the expression of EBV BZLF1. Both K-RTA and EBV BZLF1 co-localise and physically interact. During EBV lytic reactivation, BZLF1 and BRLF1 interact and synergistically activate EBV lytic reactivation. Since BZLF1 physically interact with K-RTA, this synergistic activity is inhibited by K-RTA, leading to inhibition of EBV lytic gene expression (Jiang *et al.*, 2008). The inhibition between EBV and KSHV is mutual. EBV BZLF1 also inhibits KSHV lytic gene expression (Jiang *et al.*, 2008). The mechanism of the mutual inhibition is unclear. Since BZLF1 and K-RTA both require multimerization for their function (Bu *et al.*, 2007; Sinclair, 2003), it has been suggested that the interaction between the two proteins affects the multimerization process, and therefore inhibits the function of each protein (Jiang *et al.*, 2008).

KSHV and EBV in PELs also interact *via* not only the immediate early genes, but also the latent gene LMP1. K-RTA in EBV latency cells induces LMP1 in an EBNA2 independent manner at the RNA level, K-RTA binding to the LMP1 promoter at the ISRE region and transactivates the promoter (Xu *et al.*, 2007). Overexpressing LMP1 inhibits chemically induced KSHV lytic replication in PEL. LMP1 is commonly expressed in dually infected PELs and LMP1 inhibits EBV lytic reactivation (Adler *et al.*, 2002; Prince *et al.*, 2003), by reducing the EBV BRLF1 gene *via* upregulation of NF-κB. NF-κB is known to down regulate KSHV and EBV lytic activity. Therefore, EBV inhibits KSHV lytic replication in PELs, *via* LMP1 (Xu *et al.*, 2007).

1.2. B-cell development

B-cell development is a highly regulated process, producing functional peripheral B-cell subsets from haematopoietic stem cells (HSC), in the foetal liver before birth, or bone marrow after birth. B-cell development can be summarized in several simple stages as shown in figure 1.2.1. These steps can be divided into three major stages: firstly the antigen independent stage leads to generation of mature B-cells; secondly the encounter of antigen activates the B-cells; and thirdly the antigen-driven generation of either memory B-cells or plasma cells. Each B-cell developmental stage can be distinguished by the cell surface markers and the status of immunoglobulin (Ig) genes. Each stage of development is influenced by external factors, and also regulated closely by the complex B-cell intrinsic transcription factor network. In this chapter, particular attention is paid to the final stage of this development, especially the development of plasma cells, which is associated with EBV and KSHV lytic reactivation.

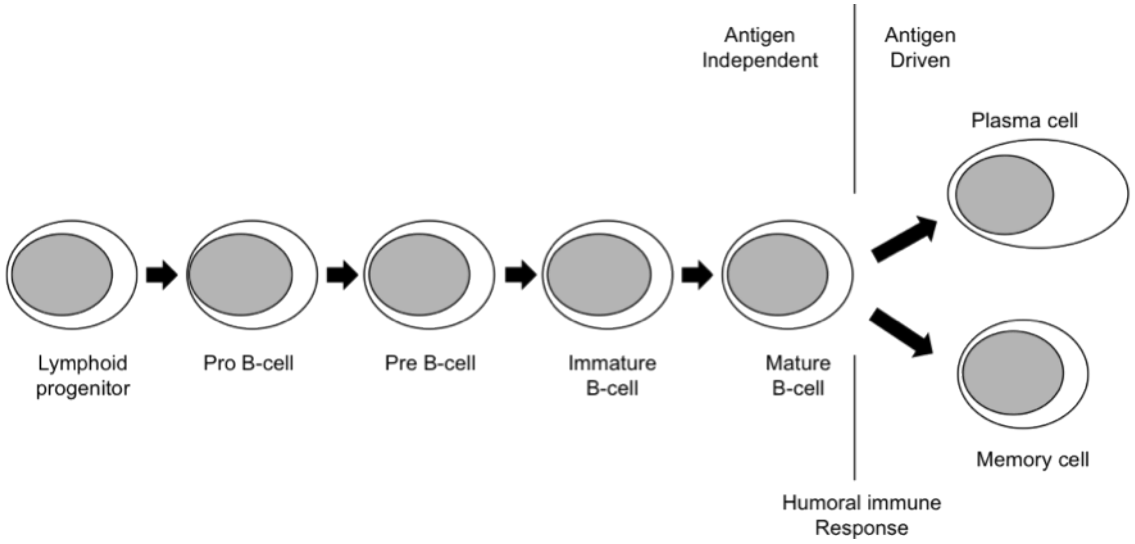


Figure 1.2.1. Schematic diagram representing the major stages of B-cell developmental starting from the lymphoid progenitor (Haematopoietic stem cells) to Mature B-cell, antigen naive cells; to antigen encounter and driven into either memory or plasma cells

1.2.1. From Haematopoietic stem cells to Mature B-cells

Human B-cells originate from the haematopoietic stem cells (HSC) in the bone marrow of adults, where the antigen independent phase occurs. The commitment of HSC to the B-cell lineage is shown by CD45 expression (Allman *et al.*, 1999), on progenitor (pro) B-cells. Pro B-cells are the first stage in the long development process driven by stimulation of various factors including stromal cell-derived growth factor 1 (SDF-1) (Nagasawa *et al.*, 1996), stem-cell factor (SCF) (McNiece *et al.*, 1991) and IL-7 (Lee *et al.*, 1988), which are secreted by surrounding cells. B-cell development is therefore dependent on the non-lymphoid stromal cells of bone marrow. Pro B-cells bind to vascular cell adhesion molecule 1 (VCAM-1) on stromal cells through the integrin very late antigen 4 (VLA-4) and also interact through cell-adhesion molecules (CAM) which promote the binding of Kit and SCF. Growth of early B-cell progenitors is stimulated by SCF which interacts with the cell-surface receptor tyrosine kinase Kit leading to the proliferation of pro-B-cell.

CD19, the major determinant of B-cells is first expressed during the pro B-cell stage. The expression of CD19 is upregulated by B-cell specific activation protein (BASP) which is encoded by Pax-5 (Adams *et al.*, 1992; Kozmik *et al.*, 1992). BASP also represses the alternative lineage choice for HSC (Nutt *et al.*, 1999) ensuring the confirmation of B-cell lineage. MicroRNAs have also been

shown to maintain the B-cell lineage at the pro B-cell stage of development (Chen *et al.*, 2004).

One of the important features of the B-cell is the variability of B-cell surface receptors (BCR) and hence the ability to recognise a wide range of antigens. The diversity of the BCR repertoire is created by the rearrangement of the immunoglobulin (Ig) chains, starting at the pro B-cell stage. The BCR consists of two heavy (H) and two light (L) chains joined by disulfide bonds, with each chain containing variable (V) and constant (C) regions. During B-cell development, both the H and L chains undergo somatic recombination resulting in the specific VDJ domains of each BCR. During the pro B-cell stage, only the H chain undergoes re-arrangement. Firstly, the diversity (D) and joining (J) gene segments of the immunoglobulin H (IgH) chain are joined together (Hardy & Hayakawa, 1991). Recombination at the V region of the DJ segment completes the IgH gene. This is referred to as VDJ rearrangement, and is catalysed by recombinase-active genes (RAG) 1 and 2 (Oettinger *et al.*, 1990) and terminal deoxynucleotide transferase (TdT) (Bassing *et al.*, 2002; Komori *et al.*, 1993). The cells then undergo positive selection for IgH rearrangement, depending on their ability to associate with protein Vλ5 and VpreB. These two proteins form a structure known as the surrogate light chain (SLC) (Cherayil & Pillai, 1991), which is similar structurally to the immunoglobulin light chain (IgL). SLCs have the ability to associate with heavy chains (Karasuyama *et al.*, 1990). The pro B-cell with the rearranged IgH forms a complex known as the pre B-cell receptor and the Igμ heavy chain associates with the SLC, allowing B-cell maturation and a proliferation burst (Hardy & Hayakawa, 1991).

Post positive selection with the pre BCR is the pre B-cell stage. With the H chain rearranged, this stage is characterised by the rearrangement of the L chain. Both κ and λ gene segments of the L chain can undergo rearrangement. However, the λ segment only rearranges if the rearrangement in the κ loci fails (Mehr *et al.*, 1999). Either only κ, or both κ and λ rearrangements result in IgL production. The expressed IgL binds to the Igμ heavy chain leading to IgM expression, resulting in an immature BCR complex with association of Igα and Igβ (Ollila & Vihinen, 2005). If the association is successful, the cells undergo negative selection and clonal deletion. Association occurs if the cells have high affinity towards self antigen (Nemazee & Burki, 1989), or the light chains of the

BCR are edited (Tiegs *et al.*, 2011) or B-cells became anergic (Goodnow *et al.*, 1989). Figure 1.2.2 shows the summary of the light and heavy chain rearrangement during both pro and pre B-cell stages.

Figure 1.2.2. Summary of events in B-cell development in bone marrow. In pro B-cells, the immunoglobulin heavy chain undergoes rearrangement and undergoes positive selection. Once cells enter the pre B-cells stage, rearrangement of immunoglobulin light chain occurs. The κ gene undergoes rearrangement before the λ genes, which only rearrange if the κ gene fails. After positive selection, the immature B-cells express surface IgM. (Adapted from (Janeway *et al.*, 2001))

The development of immature B-cells to mature B-cells occurs either in the bone marrow or *via* migration and maturation in the periphery (Allman & Pillai, 2008). Immature B-cells have a short half-life and a tendency to apoptosis post BCR engagement instead of proliferating (Allman *et al.*, 2001; Norvell *et al.*, 1995). Only 30% of the immature B-cells develop into mature B-cells (Allman *et al.*, 1993; Levine *et al.*, 2000). It is believed that the majority of the losses occur during the migration stage, whereas nearly all the migrated immature B-cells successfully develop into mature B-cells (Melchers *et al.*, 2000). During the development from immature B-cells to mature B-cells, the cells can be further divided into several distinct subsets. Originally this “transitional” B-cell stage was divided into T1 and T2 stages (Loder *et al.*, 1999) with T2 containing significant amounts of marginal zone B-cell precursors (MZPs) (Srivastava *et*

al., 2005). However, a new model has suggested there are at least three subtypes during the transitional stage, suggesting a stepwise maturation process (Allman *et al.*, 2001). The three subtypes T1, 2 and 3 all express the B-lineage precursor marker CD93/AA4 with different levels of CD21 and IgM, only the T1 subset does not express CD23 (Allman *et al.*, 2001). These three subsets of immature B-cells then undergo a series of negative and positive selections resulting in mature B-cells. Continuous BCR expression is necessary for survival of mature B-cells (Wasserman *et al.*, 1998) that also been suggested that BCR signalling plays a key role in determining the maturation into the subsets of mature B-cells (Allman & Pillai, 2008).

Naïve B-cells can be further divided into three different subsets: the follicular B2 cells, marginal zone (MZ) B-cells and the B1 cells, with the follicular B2 cells being the most common subset and the focus of chapter 1.2.2. The strength and quality of BCR signals influences the division of subsets of B-cells (Okkenhaug & Vanhaesebroeck, 2003). The follicular B2 cells circulate between the follicles in the spleen, lymph nodes and bone marrow, and are not further differentiated until they encounter an antigen. Both B1 and MZ B-cells are able to secret “natural antibodies” when they encounter an antigen even without previous antigenic experience (Martin *et al.*, 2001). Figure 1.2.3 shows the multiple stages of B-cell development.

A low affinity for BCR signalling drives the formation of MZ B-cells (Cariappa *et al.*, 2001). Both notch and NF-κB signalling are involved during the development of this subset of immature B-cells (Pillai & Cariappa, 2009). MZ B-cells circulate in humans in the same manner as the circulating memory B-cells, both cell types are also populated in the splenic marginal zone. The circulating IgM+ memory B-cells carry similar somatic mutations in the IgM genes as the MZ B-cells (Weller *et al.*, 2004). With elevated levels of a transcription factor B-lymphocyte induced maturation protein-1 (BLIMP-1), MZ B-cells differentiate into plasmablast rapidly, providing a fast immune response against blood-borne antigens (Zandvoort & Timens, 2002).

B1 cells are another subset of B-cells producing “natural antibodies” (Shapiro-Shelef & Calame, 2005). Despite being the minority subset of B-cells in adult life (Dono *et al.*, 1996), B1 cells are the majority subset in the foetal life (Bhat *et al.*, 1992). B1 cells are mainly found in the peritoneal, pleural cavities and the gut

lamina propria. Other than producing natural antibodies, B1 cells can respond to bacterial antigens, and have a BCR repertoire biased towards recognising T-cell independent type II antigens and self antigens (Martin *et al.*, 2001). B1 cells secret antibodies in the same (B-lymphocyte-induced maturation protein 1 (BLIMP-1) dependent manner as MZ B-cells, however the IgM produced by B1 cells is not mutated. B1 cells are long lived and are also capable of self-renewal (Fairfax *et al.*, 2007).

1.2.2. The Humoral immune response: B-cell activation

Antigen specific memory B-cells and antibody secreting plasma cells are essential features of the adaptive immune response, and can be derived from follicular B2 cells. Firstly the follicular B2-cells are attracted by B-lymphocyte chemokines and enter follicles of a secondary lymphoid organ (Gunn *et al.*, 1998). They develop into IgM+ and IgD+ mature B-cells, with the self-reactive B-cells which fail to develop into mature B-cells being removed under clonal deletion (Russell *et al.*, 1991). The activation of naïve and antigen specific B-cells requires two sets of signals (Lanzavecchia, 1985). Firstly, BCR signalling is required, usually initiated by either naïve or antigen specific B-cells encountering antigen that is presented by antigen presenting cells (APCs) (Gunn *et al.*, 1998; Lanzavecchia, 1985; Russell *et al.*, 1991).

The activation of B-cells is dependent on the binding strength and avidity of the antigen. This can be increased by inducing the formation of BCRs clusters and allowing successful BCR stimulation (Harwood & Batista, 2008). BCR activation recruits in expression of different signalling molecules, such as spleen tyrosine kinase (SYK) and the kinase LYN, which induces B-cells to internalise the antigen (Batista *et al.*, 2001). Once internalised the antigen is processed and presented at the cell surface as part of the MHC II complex (Lanzavecchia, 1985). This complex is recognised by the T cell receptor (TCR) of a specific population of T-helper cells that has been clonally expanded by recognition of the same antigen. The interaction between MHC II complex and the TCR of T-helper cells serve as the second signal for B-cell activation, and drives the cells to further proliferate and differentiate (Lanzavecchia, 1985).

Other than the two main signals, there are several co-stimulatory signals that are also required for the B-cell activation process, which occur between the BCR and the ligands on either T-helper cells or the APCs (Bishop & Hostager, 2001). CD40, a member of the tumour necrosis factor (TNF) receptor family, is expressed on all B-cells, with the CD40 ligand expressed on T-helper cells. CD40 and CD40 ligands are essential for class switching recombination (CSR) and germinal centre (GC) formation. OX40 and OX40 ligands are also essential for CSR (Murata *et al.*, 2000), with CD27-CD27 ligand promoting plasma cell differentiation (Jacquot *et al.*, 1997). Once activated, B-cells proliferate into plasmablasts, and can either differentiate into short live plasma cells for antibody production, or undergo a GC reaction with the help of T cells and further differentiate into either plasma cells or memory B-cells (figure 1.2.3).

Figure 1.2.3. The multiple pathways of B-cell development (adapted from (Shapiro-Shelef & Calame, 2005) by Dr. Sam Wilson). Schematic diagram represents the multiple pathways of B-cell development from the three subsets of naïve B-cells.

1.2.2.1. Germinal centre reaction

The activated proliferating B-cells form germinal centres (GC) which displace resting B-cells within the follicles of the peripheral lymphoid tissue into a surrounding mantle zone. GC consist of a dark zone of proliferating centroblasts and a light zone, formed by non-cycling centroblasts. Affinity maturation and class-switch recombination of immunoglobulin occurs during this stage of B-cell development, resulting in production of both plasma cells, that produce high affinity antibody, and long-lived memory B-cells, which provide protection against second challenge of antigens (Klein & Dalla-Favera, 2008; Schmidlin *et al.*, 2009).

Once in the germinal centre, cells undergo clonal expansion and somatic hypermutation (SHM) (Berek *et al.*, 1991; Jacob *et al.*, 1991), which modify the genes encoding the variable regions of immunoglobulin heavy and light chains. SHM, facilitated by activation induced cytidine deaminase (AID) (Muramatsu *et al.*, 2000; Revy *et al.*, 2000; Stavnezer & Schrader, 2006), introduces substitutions, deletions and insertions into the VDJ recombination sites of Ig (Goossens *et al.*, 1998). The affinity for specific antigen alters due to the mutation at the BCR variable region, with approximately one mutation introduced per cell division (McKean *et al.*, 1984). GC B-cells with high affinity antigen receptors outcompete the low affinity GC B-cells, differentiate into non-dividing centrocytes and migrate into the light zone of germinal centre. The GC-B-cells with low affinity antigen receptors undergo apoptosis (Takahashi *et al.*, 1999). The light zone also contains follicular dendritic cells (FDCs), T-cells and macrophages that help select GC B-cells with high affinity towards the antigen. Class switching recombination (CSR) may occur in the light zone at this stage (Toellner *et al.*, 1996), allowing B-cells to change the surface immunoglobulin from IgM to IgD, A, G or E. The different IgH constant regions are arranged sequentially in the human genome; with CSR inducing non-homologous end-joining recombination between the switch regions upstream of the heavy chain constant regions, resulting in different IgH gene expression without modifying the variable region (Stavnezer & Schrader, 2006). This process is linked to SHM and requires AID to increase antibody diversity (Hackney *et al.*, 2009; Longerich *et al.*, 2006; Ta *et al.*, 2003). Uracil-N-glycosidase (UNG) is also required for CSR (Rada *et al.*, 2002).

The GC B-cells travel between the light and the dark zone of the GC, undergoing further clonal expansion, SHM and CSR, a process refered to as affinity maturation. Eventually, the centrocytes that survive all the GC reactions migrate from the GC and further differentiate into memory B-cells, or plasmablasts (pre antigen-produced plasma cell) (Allen *et al.*, 2004). The signal required for centrocytes to exit the GC is unclear, as are the signals required for GC B-cell differentiation. Previous studies indicated that the production of memory B-cells by the germinal centre reaction is random (Blink *et al.*, 2005). High affinity BCR however drives the post GC-B-cells to differentiate into plasma cells (Phan *et al.*, 2006), which is controlled and regulated by transcriptional factors.

The transcription factor B-cell lymphoma 6 (BCL-6) is present at high levels in GC B-cells (Cattoretti *et al.*, 1995) and is required for GC formation (Dent *et al.*, 1997). BCL-6 acts as a transcriptional repressor and downregulates the genes involved in growth suppression and apoptosis. This allows cells in GC to proliferate and undergo SHM and CSR (Parekh *et al.*, 2007; Shaffer *et al.*, 2000). BCL-6 also suppresses the gene required for T-cell dependent B-cell activation, preventing immature differentiation of B-cells (Shaffer *et al.*, 2000). The most important feature of BCL-6 is the suppression of B-lymphocyte-induced maturation protein 1 (BLIMP-1) expression which allows the GC reaction to continue before differentiation into plasma cells (Fearon *et al.*, 2002; Shaffer *et al.*, 2000) BCL-6 represses positive-regulatory-domain-containing-1 (Prdm1) which encodes BLIMP-1. See figure 1.2.4. for detail. The metastasis-associated 1 family member 3 (MTA3) is required for BCL-6 mediated repression of BLIMP-1.

However, with continuous BCL-6 expression, GC B-cells cannot differentiate into plasma cells. BCR signalling has been shown to suppress BCL-6 expression *via* proteasome mediated degradation of BCL-6. This alone is not sufficient to induce GC B-cells into plasma cell differentiation and other signals may be required for this process (Schliephake & Schimpl, 1996; Shapiro-Shelef & Calame, 2005). CD40 signalling is able to down regulate BCL-6 (Panagopoulos *et al.*, 2004), suggesting an involvement of T-cells in the final stages of plasma differentiation.

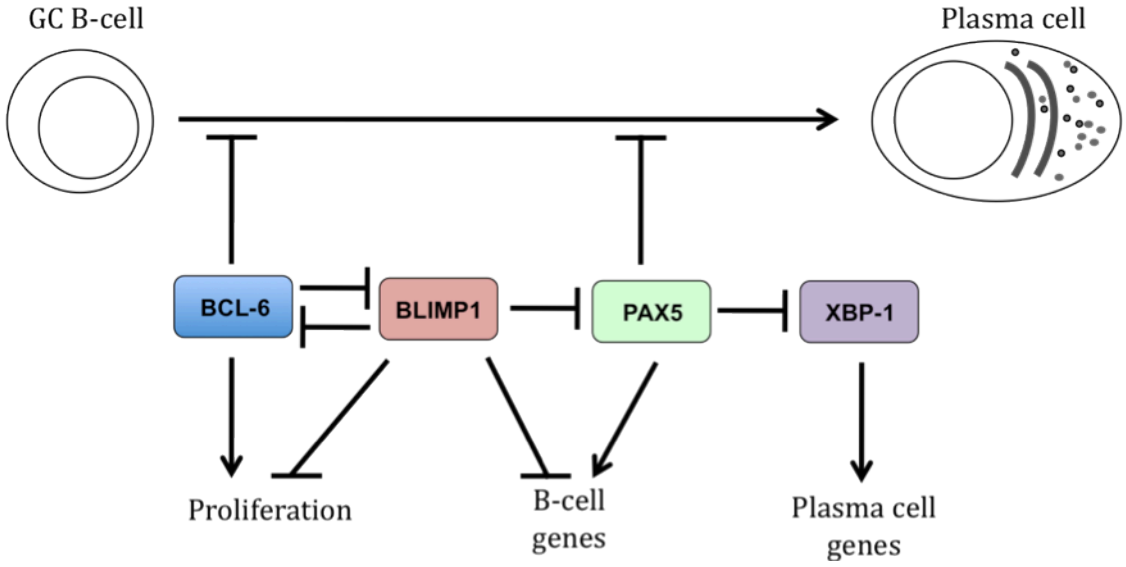


Figure 1.2.4. Transcription factors involved in regulating the terminal differentiation into plasma cells from GC B-cells.

1.2.3. Memory B-cells

Memory B-cells are long-lived non-dividing cells, which circulate either in the blood (Benner *et al.*, 1977) or colonise the MZ of the spleen (Liu *et al.*, 1988). Memory B-cells can rapidly proliferate and differentiate into plasma cells at the second encounter with the original antigen (Arpin *et al.*, 1997; Mitchell *et al.*, 1972) or respond to polyclonal stimulation through pattern recognition and T-cell help (Bernasconi *et al.*, 2003; Bernasconi *et al.*, 2002). Interestingly, naïve B-cell do not respond to this signal (Bernasconi *et al.*, 2003). This allows memory B-cells to maintain the memory pool in the absence of an antigen.

All memory B-cells are IgD^-, with CD27 being identified as the memory B-cell marker (Klein *et al.*, 1998), and can be divided into subsets according to the Ig isotype expressed. The locations of memory B-cell subsets are also restricted to facilitate a good immune response, for example memory B-cells expressing inhibitory receptor Fc receptor like 4 (FCRL4) are restricted to mucosal tissues and amesenteric lymph nodes (Ehrhardt *et al.*, 2005). The memory B-cell is long-lived and can persist for the lifetime of a host, through high expression of anti-apoptotic factors such as Bcl-2, A1 and Mcl-1 compared with naïve B-cells (Good *et al.*, 2009). However the full mechanism for memory persistence remains unclear.

Although BCL-6 is an important factor for preventing GC B-cells from differentiating into plasma cell and memory B-cells, Toyama *et al*. (2002) showed that memory B-cells are still able to be produced in BCL-6 deficient mice, despite them being unable to form GC. Nevertheless, over-expression of BCL-6 leads to maintenance GC B-cells and prevents differentiation into memory cells (Kuo *et al.*, 2007). Therefore; the differentiation from GC B-cells to memory B-cells involves other transcription factors. Post GC memory B-cells can be sub-divided into two subsets: the CD45+ memory B-cells, which are highly proliferative (Bleesing & Fleisher, 2003), and the CD40- pre-plasma memory B-cells, which are less proliferative but can differentiate into plasma cells quickly (Driver *et al.*, 2001; McHeyzer-Williams *et al.*, 2000). B-cells deficient in BLIMP-1 expression, resulting in high-level expression of BCL-6, can only develop into post GC memory cells, but not pre plasma memory cells. This suggests a linear memory cell development with post GC memory B-cells upstream of pre-plasma memory B-cells (Shapiro-Shelef *et al.*, 2003). However, further investigation is required to support this model.

1.2.4. Plasma cells

Plasma cells are non-dividing effector cells of the B-cell lineage. B-cells eventually terminally differentiate into plasma cells and are able to syntheses and secrets large amounts of immunoglobulin against a specific antigen, acting as an essential feature of both innate and adaptive immunity. Plasma cells can be differentiated from many different B-cell stages, including follicular B2 cells, GC B-cells, and memory B-cells, depending on the nature of the antigen and the location of the encounter. This process is tightly regulated by many transcription factors as shown in figure 1.2.4.

B-lymphocyte-induced maturation protein 1 (BLIMP-1) is a five-zinc-finger-containing protein that was first identified as a repressor of the human interferon-β (IFN-β) promoter (Keller & Maniatis, 1991) and was the identified as an induced transcript when BCL-1 lymphoma cells differentiate into antibody secreting cells after cytokine treatment (Turner *et al.*, 1994). Exogenous expression of BLIMP-1 in mature B-cells or cell lines can induce plasma cell differentiation. BLIMP-1 knock out mice do not develop plasma cells (Shapiro-

Shelef *et al.*, 2003; Turner *et al.*, 1994). BLIMP-1 regulates three transcription factors (Shaffer *et al.*, 2002); BLIMP-1 directly down regulates paired box protein 5 (PAX5) (Lin *et al.*, 2002) and BCL-6 (Shaffer *et al.*, 2002), and indirectly activates XBP-1 and BLIMP-1. BLIMP-1 acts upstream of XBP-1 which is required for terminal differentiation. However, if BLIMP-1 is deficient in cells, over-expression of XBP-1 alone cannot induce plasma cell differentiation (Shapiro-Shelef *et al.*, 2003). Many of the genes required for regulating terminal differentiation of plasma cells are regulated by both BLIMP-1 and XBP-1 proteins (Shaffer *et al.*, 2004).

PAX5 is also known as B-cell-lineage-specific activator protein (BSAP), and regulates many genes involved in Ig transcription, BCR signalling and B-cell identity, *via* binding to DNA by a bipartite paired domain, and can function as either a transcriptional activator or a repressor (Nutt *et al.*, 2001). Similar to BCL-6, PAX5 represses GC B-cell terminal differentiation of plasma cells. One of the main functions of PAX5 is the direct repression of XBP-1 transcription (Reimold *et al.*, 1996). PAX5 is also able to down regulate the expression of IgH, IgL and the J chain. Since expression of Ig can also induce XBP-1 expression, by repressing the expression of the Ig, PAX5 also indirectly blocks XBP-1 expression.

1.2.4.1. Unfolded protein response

The terminal differentiation of B-cells into plasma cells occurs through a transcription factor mediated development chain. During this process, the B-cells expand their secretary apparatus (Shaffer *et al.*, 2004; Shapiro-Shelef & Calame, 2005) ready for large scale Ig secretion. This process involves the activation of transcription factors associated with the unfolded protein response (UPR); an inducible stress response that is able to increase the capacity for protein folding in the endoplasmic reticulum (ER) (Schroder & Kaufman, 2005) whilst decreasing synthesis of proteins and decreasing cell cycle arrest and apoptosis (Gass *et al.*, 2004). Some functions during the UPR such as ER stress have been related to many diseases, such as diabetes, inflammation, autoimmunity and neurodegenerative disorders; whereas other functions of UPR, such as decreasing protein synthesis, have not. Therefore it has been

suggested that terminal differentiation of plasma cells is through a "physiological" UPR, rather than the "stress induced" UPR (Blond-Elguindi *et al.*, 1993; Gass *et al.*, 2002; Shaffer *et al.*, 2004). The UPR in mammalian cells is more complex than yeast, where it was originally discovered, and involves three branches of signalling, initiated by three different ER integral membrane proteins: activating transcription factor-6 (ATF6), protein kinase RNA (PKR)-like ER kinase (PERK) and IRE-1α. The signalling pathways between these three branches cross talk and contain various positive and negative feedbacks. The production of unfolded proteins in the ER initiates the UPR and it is sensed and mediated through BiP (also known as GRP78), an IgH binding protein (Haas & Wabl, 1983). BiP binds to the short hydrophobic motifs that reside in the interior of soluble proteins in the ER, but with low affinity (Blond-Elguindi *et al.*, 1993; Flynn *et al.*, 1991). This enables BiP to be recruited by the unfolded protein and removed from its usual binding partners, such as PERK, IRE1 or ATF6. This initiates the UPR. Figure 1.2.5 shows the detail of the UPR in mammalian cells.

PERK and IRE-1α both contain transmembrane serine/threonine kinase cytoplasmic domains (Bertolotti *et al.*, 2000). Following dissociation with BiP, both PERK and IRE-1α undergo homodimerisation and autophosphorylation. PERK then phosphorylates eularyotic initiation factor-2α (eIF-2α) (Harding *et al.*, 1999) causing translation attenuation, thereby preventing further accumulation of proteins in the ER. However, certain protein translation such as ATF4, which is required for recovering from cellular stress is active and requires phosphorylation of eIF-2α (Harding *et al.*, 2000; Harding *et al.*, 2003; Lu *et al.*, 2004). Expression of ATF4 upregulates genes involved in amino acid synthesis and also activates C/EBP homologous protein (CHOP), which is a pro-apoptotic transcription factor (Harding *et al.*, 2000). CHOP expression induces the expression of growth arrest and DNA damage inducible protein 34 (GADD34), leading to dephosphorylation of eIF-2α *via* protein phosphatase 1 (PP1). CHOP expression is also induced by ATF6. CHOP also inhibits the expression of Bcl-2 (Zinszner *et al.*, 1998); Bcl-2 is important in switching the signalling of UPR from pro-survival to apoptotic (Rutkowski *et al.*, 2006). Hence ATF4 acts as a negative feedback regulator of PERK mediated inhibition of translation, relieving the cells from the translational cell repression, allowing cells to express proteins that are required for cellular recovery (Marciniak *et al.*, 2004).

Figure 1.2.5. ER stress and activation of the UPR signalling pathways. Accumulation of misfolded or unfolded protein aggregates in the ER lumen, a condition known as ER stress, leads to activation of three ER transmembrane proteins, PERK, IRE1, and ATF6. GRP78, a ubiquitous ER chaperone that is normally bound to these ER stress sensors and keeps them inactive, dissociates from them in order to assist with the folding of proteins in the ER lumen. However, this dissociation leads to activation of the 3 UPR pathways. (1) PERK homodimerization and autophosphorylation results in the subsequent phosphorylation of the *α* subunit of eIF2 which by inhibiting global protein synthesis reduces the ER protein load. ATF4 expression, however, increases upon eIF2*α* phosphorylation which translocates to the nucleus allowing for transcription of UPR target genes by binding to the UPR response element (UPRE). These genes include CHOP, a pro apoptotic transcription factor that results in cell death if ER stress conditions persist, and GADD34, which acts as a negative regulator of the PERK pathway by dephosphorylating eIF2*α*. (2) IRE1 is activated in a similar manner to PERK by homodimerization and autophosphorylation. Additionally, interaction of misfolded or unfolded proteins with the luminal domain of IRE1 can also further promote its activation. *XBP1* mRNA is an IRE1 substrate that undergoes splicing to produce *XBP1s*, encoding a transcription factor that can lead to upregulation of ER chaperones and other UPR target genes. (3) ATF6 activation leads to its translocation to the Golgi where it is sequentially cleaved by site 1 and site 2 proteases. This leads to the release of the N-terminal ATF6 fragment which translocates to the nucleus, binds to the ER stress response element (ERSE) thereby activating UPR target genes. (Basseri & Austin, 2012)

ATF6 is comprised of two transmembrane bZIP transcription factors, ATF6α and ATF6β (Haze *et al.*, 1999). Under normal conditions, these two factors are held together in the ER with BiP. After dissociation from BiP, these two factors migrate to the Golgi and are cleaved by site 1 and site 2 proteases, releasing

the N-terminal domain of ATF6 (Ye *et al.*, 2000). ATF6β has a minimal role in the UPR (Thuerauf *et al.*, 2004); however ATF6α entry into the nucleus activates many UPR target genes, such as XBP-1, CHOP and ER chaperons. With both ATF6 and XBP-1 being bZIP transcription factors, formation of either homodimers or hetrodimers of these two proteins are possible (Newman & Keating, 2003). ATF6 activates the genes *via* inducing transcription from the promoters containing ER stress responsive elements (ERSEs), resulting in expression of BiP and protein disulphide isomerase (PDI), which increases the folding capacity of ER, and activates the genes involved in ER associated protein degradation (ERAD) (Okada *et al.*, 2002; Yamamoto *et al.*, 2007). ERAD allows unfolded proteins to be translocated from the ER lumen into the cytosol, and degraded by the proteasome, reducing the amount of unfolded protein in the ER. Other than promoting unfolded protein degradation and inducing XBP-1 expression, ATF6 is also important in ER expansion and lipid biosynthesis (Bommiasamy *et al.*, 2009). As well as sensing ER stress through BiP disassociation, ATF6 can also sense ER stress *via* alternative pathway. ATF6 contains three glycosylation sites; and once glycosylated, it promotes interaction with calreticulin. This interaction allows ATF6 to anchor in the ER, whereas the unglycosylated ATF6 migrates to the Golgi (Hong *et al.*, 2004) and initiates UPR. This alternative activation pathway of ATF6 could allow UPR to be activated by different stress signals.

IRE-1α also undergoes homodimerization and autophosphorylation following BiP dissociation, which activates both kinase and endoribonuclease activity of IRE-1α (Hosoi *et al.*, 2010). Unfolded proteins may also directly bind to IRE-1α and induce homodimerisation and autophosphrylation (Credle *et al.*, 2005; Gardner & Walter, 2011; Kimata *et al.*, 2007) without dissociation from BiP. IRE-1α can then bind to tumour-necrosis factor (TNF)-receptor-associated factor 2 (TRAF2) which then activates JNK signalling *via* apoptosis signal regulating kinase-1 (ASK-1) (Nishitoh *et al.*, 2002; Urano *et al.*, 2000). Activation of JNK results in activation of autophagy, which removes damaged organelles and aggregated proteins (Bernales *et al.*, 2006; Ogata *et al.*, 2006), and apoptosis of the cells that are unable to recover from ER stress (Yoneda *et al.*, 2001). IRE-1α in UPR is linked to apoptosis through several members of Bcl-2 family such as BAX, BAK and “BH3 only”, there have been shown to interact with IRE-1α,

resulting in apoptosis (Hetz *et al.*, 2006; Klee *et al.*, 2009). IRE-1α expression also induces mRNA degradation *via* the endoribonuclease domain, reliving the pressure of ER. The function of ATF6, IRE-1α and PERK highlights two proposes of UPR: firstly attempted recovery and secondly, if homeostasis cannot be restored, induce apoptosis.

1.2.4.2. XBP-1 in plasma cell differentiation

The transcription factor X-box binding protein 1 (XBP-1) is a bZIP transcription factor identified by its ability to bind to the X-box, which is a conserved transcriptional element in the human leukocyte antigen (HLA)-DRα promoter (Liou *et al.*, 1990). Expression of XBP-1s is essential for terminal differentiation of plasma cells (Reimold *et al.*, 2001). Over expressing XBP-1 in post GC B-cells or plasmablasts drives the cells to express transcriptional profiles of plasma cells (Shaffer *et al.*, 2004; Wilson *et al.*, 2007). Expression of XBP-1 mRNA is induced by ATF6 (Yoshida *et al.*, 2000), IL-4 (Iwakoshi *et al.*, 2003) and BLIMP-1-dependent repression of PAX5 expression (Shapiro-Shelef *et al.*, 2003). XBP-1 is expressed retaining a 26 nucleotides (nt) intron (XBP-1u, unspliced) that results in a frame shift preventing the translation of an active transcription factor. Under ER stress and B-cell terminal differentiation, the 26nt intron of XBP-1u mRNA is removed to allow the translation of the full length stable and active form of XBP-1s (XBP-1 spliced) (Calfon *et al.*, 2002). The splicing event of XBP-1 mRNA is facilitated by IRE-1α (Calfon *et al.*, 2002) and is dependent on the increased IgH expression during the UPR (Iwakoshi *et al.*, 2003). XBP-1 induces expression of $p58^{IPK}$, which negatively regulates PERK expression, indicating further cross talk between the UPR pathways (Yan *et al.*, 2002). The inactivated XBP-1u acts as dominant-negative regulator of XBP-1s by occupying the binding sequence of target promoters (Lee *et al.*, 2003). Accumulation of XBP-1u during ER stress recovery also blocks XBP-1s nuclear entry (Yoshida *et al.*, 2006). Hence XBP-1u acts as negative regulator of XBP-1s and degradation of XBP-1u is essential for XBP-1s activity.

In B-cells, XBP-1 induces the gene expression involved in targeting of proteins to ER, translocation into the ER, folding of proteins, degradation of misfolded proteins, glycosylation of proteins, trafficking between ER and the Golgi, trafficking in the ER pathway, and targeting secretary vesicles to the plasma

membrane (Shaffer *et al.*, 2004). XBP-1 also induces the expansion of mitochondria mass, ER, cell size, lysosomal content, ribosomal content, and total protein synthesis (Shaffer *et al.*, 2004); suggesting the main function of XBP-1 is to facilitate the development of secretory apparatus in plasma cell differentiation. Recent studies have indicated that XBP-1 activation is differentiation dependent and occurs prior to UPR during B-cell terminal differentiation (Hu *et al.*, 2009; Todd *et al.*, 2009).

1.2.5. EBV control of B-cell development

EBV is known to be able to manipulate B-cell development, and can transform B-cells into immortal LCLs *in vitro*. Despite life-long EBV latent infection in the memory B-cells, EBV actively infects naïve B-cells *in vivo* and manipulates cellular factors by mimicking certain aspects of B-cell differentiation thereby driving the infected cell towards a memory cell phenotype (Souza *et al.*, 2005; Thorley-Lawson & Babcock, 1999). Figure 1.2.6 shows a side-by-side comparison model of the EBV manipulation of B-cell differentiation after infection of naïve B-cells. EBV establishes different latency programs throughout the differentiation in different type of B-cell, which is shown in table 1.1.2.

Figure 1.2.6. B-cell development and EBV (adapted from (Thorley-Lawson, 2005) by Dr. Sam Wilson). A diagrammatic representation of normal and EBV-driven B-cell differentiations. EBV gene expression programs are highlighted in grey.

1.2.5.1. Infection of resting naïve B-cells

EBV is transmitted through saliva and infects IgD positive naïve B-cells in the nasopharyngeal lymphoid tissue (Joseph *et al.*, 2000). Upon immediate infection, EBV enters an initial burst of lytic replication. Virus particles are

produced during this period. This early lytic cycle is important and contributes to the establishment of latency and transformation of B-cells (Halder *et al.*, 2009). Naïve B-cells express Pax-5 which binds to and activates the Wp of EBV which induce the expression of EBNA2 (Tierney *et al.*, 2000a). EBNA2 then activates other viral genes, including the rest of the EBNA family proteins and the LMP family proteins (Abbot *et al.*, 1990), allowing EBV to establish the latency III program (growth program). This mimics the affect of antigen encounter and activates the B-cells (Wang *et al.*, 1987), driving them to follicles where they undergo germinal centre like reactions and the EBV gene expression profile is switched to latency II.

Interestingly, despite the importance of LMP proteins in transforming B-cells (see section 1.2.5.2.), the number of W repeats in EBV genome can affect the ability of virus transformation. EBV with lower numbers of W repeats is reported to have impaired ability in B-cell transformation compared to a virus with more W repeats (Tierney *et al.*, 2007). EBV with one W repeat is virtually incapable of transforming B-cells; the ability to affect transformation reaches a peak at 5 copies of W repeats (Tierney *et al.*, 2011). The number of W repeats directly affects the function of Wp, which encodes EBNA2, EBNALP and BHRF1 at the Wp-initiated transcript. It has been suggested that the difference in transformation ability is due to the suboptimal levels of Wp transcription from the incoming virus genome, and therefore suboptimal levels of EBNA2 and EBNALP, two key transactivators of both viral and cellular gene expression. EBV impaired transformation ability due to small number of W repeats can be rescued by over expressing both full length EBNA2 and EBNALP proteins (Tierney *et al.*, 2011). However, other studies have also shown that EBNALP with two repeat domains could efficiently co-activate the Cp and LMP1 promoter (Rickinson & Kieff, 2006; Shannon-Lowe *et al.*, 2009).

1.2.5.2. Germinal centre B-cells

Latency program II is the program expressed in the GC cells and is also known as the default program, with only three viral genes expressed: EBNA1, LMP1 and LMP2. It is suggested that the switch between Latency III and II programs may be caused by EBNA2 hyperphosphorylation (Yue *et al.*, 2004) which

downregulates EBNA2, although further evidence is required for the hypothesis. In EBV free B-cell development, GC B-cells undergo a series of somatic hypermutation and class switching recombination which are signalled by antigen encounter and T helper cells. In EBV positive B-cells, LMP1 and 2 mimic these signals in the GC reaction.

LMP1 mimics the signals generated through CD40-CD40 ligands provided by T-helper cells (Kilger *et al.*, 1998; Zimber-Strobl *et al.*, 1996), *via* ligand independent recruitment of tumour necrosis factor (TNF) receptor-associated factors (TRAFs). LMP1 also upregulates the expression of BAFF and induces ligand-independent CSR (Bleesing & Fleisher, 2003). LMP2A mimics BCR signalling through association with the tyrosine kinase Lyn (Burkhardt *et al.*, 1992), and has been shown to be able to drive B-cells into GC formation independent of antigen in transgenic mice (Casola *et al.*, 2004). LMP1 also inhibits expression of BCL-6, which signals the B-cells to leave the germinal centre and enter the memory cell pool.

1.2.5.3. Peripheral blood memory B-cells

EBV positive post GC B-cells can either differentiate into memory B-cells or plasma cells after exiting the GC reaction. The differentiation into memory B-cells is associated with EBV achieving life-long persistency. The circulating EBV positive B-cells express latency I/0 profiles, with minimal viral protein expression. Differentiation into plasma cells is associated with EBV reactivation. Studies have shown EBV could be reactivated by XBP-1s expression alone or in combination with protein kinase D (PKD) during terminal differentiation into plasma cells (Bhende *et al.*, 2007; Laichalk & Thorley-Lawson, 2005). Other studies suggested ER stress induces EBV into lytic reactivation (Taylor *et al.*, 2011). More details on EBV reactivation and plasma cell differentiation will be discussed in chapter 3.

1.3. Gamma Herpesvirus Associated tumours

1.3.1. B-cell tumourigenesis

Different B-cell tumours have been associated with B-cells arrested at different stages of the developmental process. Fundamental understanding of each B-cell development stage will help future studies of these tumours. Since EBV is capable of transforming and affecting gene expression in B-cell development, it has also been suggested that EBV infection may contribute towards formation of B-cell tumours. A hypothesis suggested is that EBV negative lymphomas lose EBV after the lymphoma has formed. Sporadic BL without detectable EBNA1 had been found to contain detectable partial EBV genomes (Razzouk *et al.*, 1996). It has been suggested that EBV had an initiating role in tumorigenesis. Despite losing EBV later on in the process, the cells had acquired other mutations during the progression of malignancy, and these mutations compensated for the lost of EBV. This therefore leads to lymphomas being clinically presented as EBV negative (Vereide & Sugden, 2011). Here we review some of the B-cell malignancies that have been associated with EBV or herpesvirus infection, with the exception of NPC, which is of epithelial origin but is associated with EBV infection.

1.3.2. Primary Effusion Lymphoma

Primary effusion lymphoma (PEL) is a neoplasm of large B-cells usually presenting as serous effusions without detectable tumour masses, usually in the cavity of human body, and was thus initially referred to as body cavity based lymphoma (BCBL) (Knowles *et al.*, 1989). However, cases of isolated solid lymph-node associated PEL have been discovered since (Chadburn *et al.*, 2004). With KSHV being associated with BCBL (Cesarman *et al.*, 1995b), PEL was redefined as a KSHV positive lymphoma (Nador *et al.*, 1996). 80% of PEL are also EBV positive. PEL accounts for about 3% of the AIDS-related non-Hodgkin's lymphomas (Knowles, 2003); the majority of PEL patients are HIV positive and immune-suppressed (Komanduri *et al.*, 1996), however PEL can also occur in HIV negative individuals (Nador *et al.*, 1996). The only difference between PEL in HIV positive and negative individuals is the latter tends to occur later on in life (Carbone *et al.*, 1996; Nador *et al.*, 1996).

Despite being of B-cell origin, PEL does not express CD19 on the cell surface. PEL expresses CD138, CD30, CD38, CD45 and CD71, which is similar in immune phenotype to the cycling plasmablast. Most PELs have hypermutated immunoglobulin (Ig) gene, which suggests that they have passed through the germinal centre reaction (Fais *et al.*, 1999; Matolcsy *et al.*, 1998). The gene expression profiles of PELs indicate that they are arrested in the stage of a plasmablast in B-cell development (Jenner *et al.*, 2003; Klein *et al.*, 2003); however, PELs do not express Ig on the cell surface (Nador *et al.*, 1996). CD138 expression is typically seen in either pre-B-cell or plasma cells and no other lymphomatous effusion (Gaidano *et al.*, 1997), and can therefore be used as a marker to diagnose PEL.

1.3.3. Burkitt's Lymphoma

Burkitt's Lymphoma (BL) is named after Dr. Denis Parsons Burkitt who first described the disease in equatorial Africa; it is also the tumour where EBV was first discovered. BL is a highly aggressive lymphoma that often presents at extranodal sites, or as acute leukaemia. It contains monomorphic medium-sized B-cells with basophilic cytoplasm and numerous mitotic figures. BL shows the same pattern of immunoglobulin gene hypermutation as germinal centre B-cells and memory B-cells (Klein & Zachau, 1995), but has the cellular phenotype of a germinal centre B-cell (Gregory *et al.*, 1987). EBV often establishes latency type I program in BL cell lines.

BL can be divided into three variants: endemic, sporadic and immunodeficiency associated variants. The endemic variant of BL occurs in equatorial Africa and is associated with EBV. The sporadic variant of BL occurs worldwide, and only about 15% of sporadic BL are EBV positive (Rickinson & Kieff, 2001); however, it is suspected that EBV infection contributes to the formation of sporadic BL but the cells then lose the infection later on in the tumour formation process (Razzouk *et al.*, 1996; Vereide & Sugden, 2011). AIDS is often associated with BL however only 30-40% of AIDS related BLs are EBV positive (Diebold, 2001).

Endemic BL (eBL), with or without EBV infection, has been long associated with holoendemic *Plasmodium falciparum* malaria geographically (Burkitt, 1958; Dalldorf *et al.*, 1964) and also has a molecular link to holoendemic malaria

(Chene *et al.*, 2007). Previous studies have shown the cysteine-rich interdomain region 1α (CIDR1α) of the malaria protein *P falciparum* erythrocyte membrane protein 1 (PfEMP1), which is expressed at the surface of parasite positive red blood cells, activate B-cells, especially the membrane B-cells, where EBV remains latent (Chene *et al.*, 2007; Donati *et al.*, 2006). However, the interactions between EBV infection, risk of BL and malaria infection remain unclear and require future study.

C-myc is a transcription factor that is involved in many cellular progresses such as cell growth, proliferation and apoptosis. The common feature of all BL is the c-myc chromosomal translocation occurs most frequently at chromosome 8q24 to the Ig heavy chain region on chromosome 14q32, which deregulates the expression of c-myc oncogene (Klein, 1983; Leder, 1985; Manolov & Manolova, 1972). This accounts for about 85% of the c-myc translocation in BL. The remaining 15% is made up of either translocations at chromosome 2q8 or 8q22, at the kappa or lamda IgL region. The translocation leads to high level of c-myc expression (Klein *et al.*, 1968). Li *et al.* (2003) have shown that c-myc binds and is involved in the regulation of large number of gene promoters in BL and promoties tumorigenesis.

1.3.4. Post-transplant lymphoproliferate disease (PTLD)

PTLD is a lymphoid proliferation or lymphoma that develops as a consequence of immunosuppression in a recipient of a solid organ or bone marrow allograft. PTLD refers not to a single disease but to a wide range of lymphoproliferations, ranging from early EBV driven polyclonal proliferations resembling infectious mononucleosis (IM), to EBV positive or negative lymphomas of predominantly B-cells, or less commonly, T-cells. PTLD can affect nearly all organ system and is able to form a heterogeneous group of tumours, ranging from B-cell hyperplasia to immunoblastic lymphoma. Approximately 90% of the PTLD are B-cell origin, and 90-95% contain EBV (Harris *et al.*, 2001). However, the association of PTLD with EBV is dependent on many factors such as the type of organ allograft, the patient population (adult or paediatric), and the immunosuppressant regiment (Nalesnik, 1996; 1998).

The frequency of PTLD varies with the type of graft. It is estimated that PTLD accounts for 2% frequency in solid organ transplant recipients (Nalesnik, 1996; 1998). With the lowest incident in renal transplant, which is around 1%, patients receiving lung, heart and liver transplant have higher incident frequency of developing PTLD, which is around 5% (Harris *et al.*, 2001). Paediatric patients have higher frequency of developing PTLD than adults (Dror *et al.*, 1999; Harmon & Dharnidharka, 1999; Ho *et al.*, 1988; Pickhardt *et al.*, 1998). The general frequency of children under 15 developing PTLD post solid organ transplant is about 8% (Pickhardt *et al.*, 2000). The frequency of paediatric PTLD post liver transplant is 4-15%, lung and heart transplant is 7.7-19.5% and 1-8% in kidney transplant (Boyle *et al.*, 1997; Cox *et al.*, 1995; Dror *et al.*, 1999; Harmon & Dharnidharka, 1999; Malatack *et al.*, 1991; Newell *et al.*, 1996; Pickhardt *et al.*, 1998; Younes *et al.*, 1999). The majority of the EBV in PTLD after solid organ transplant is of host origin, with a minority (less than 10%) of donor origin. However, this also varies depending on the type of grafts. Majority of the patients receiving bone marrow transplant developed PTLD of donor origin EBV. Increased levels of immunosuppression allows greater EBV expansion; therefore a high level of immunosuppression is another risk factors for developing PTLD, especially when treating rejections (Cox *et al.*, 1995; Harmon & Dharnidharka, 1999; Newell *et al.*, 1996).

Since PTLD represents a large number of lymphoproliferative diseases, the America Society of Haematopathology has outlined three general categories for PTLD according to the morphology (Harris *et al.*, 1997). Early lesions: which include plasma cell hyperpasia and infectious-mononucleosis-like-PTLD. This type of PTLD shows preservation of the nodal architecture (Harris *et al.*, 2001; Pickhardt *et al.*, 2000). Polymorphic PTLD are defined as destructive lesions composed of immnuoblasts, plasma cells, and intermediate sized lymphoid cells. This type of PTLD also shows destruction of the nodal architecture and locally invades the involving organs (Pickhardt *et al.*, 2000). The third type of PTLD is referred to as monomorphic PTLD, which is characterised by nodal architecture effacement and/or invasive tumour growth in extra nodal sites with confluent sheets of transformed cells (Harris *et al.*, 2001). This type of PTLD can be further divided according to histopathological features by using standard lymphoma nomenclature, such as diffused large B-cell lymphoma, Burkitt's lymphoma or multiple myeloma.

1.3.5. Hodgkin's lymphoma

Hodgkin's Lymphoma can be further divided into Nodular lymphocyte predominant Hodgkin lymphoma (NLPHL) and classic Hodgkin's lymphoma (CHL), depending on the morphology, immunophenotype, immunoglobulin transcription and the composition of the cellular background. EBV infection can only be found in the CHL.

CHL accounts for 95% of the HLs, and has a bimodal age curve, the first between ages 15-35 and the second one later on in life. It is a monoclonal lymphoid neoplasm composed of mononuclear Hodgkin cells and multinucleated Reed-Sternberg (HRS) cells. It resides in an infiltrate containing a variable mixture of non-neoplasmic small lymphocytes, plasma cells, fibroblasts and collagen fibres. EBV can only be found in a population of CHL, with HIV often being associated with EBV positive CHLs. Previous reports have associated patients with infections mononucleosis with higher incidence of CHL (Mueller, 1999). EBV positive CHL expresses the latency type II program (Stein, 2001).

1.3.6. Nasopharyngeal Carcinoma

Nasopharyngeal carcinoma (NPC) is a tumour of epithelial cells in the nasopharynx. As discussed earlier in this chapter, it is more common in Asia. NPC, according to the classification by the World Health Organisation (WHO), can be divided into three types. Type I is referred to as the squamous cell carcinoma, type 2 is the Keratinizing undifferentiated carcinoma, and type III is the nonkeratinizing undifferentiated carcinoma. Type III accounts for 80% of the NPC cases and is strongly associated with EBV. There are several aetiological factors associated with NPC, firstly genetic susceptibility, secondly, an early exposure to chemical carcinogens and finally, association with EBV latent infection (Busson *et al.*, 2004; Chan & Lo, 2002). EBV in NPC arises from a single EBV infected progenitor cell (Raab-Traub & Flynn, 1986). The progression of NPC requires accumulation of many changes, either genetic or epigenetic in the host genome (Lo & Huang, 2002).

EBV is in latency II within the NPC cells, with expression of high level of LMP2A (Brooks *et al.*, 1992), and LMP1 (Chen *et al.*, 1995). Low level of lytic replication

can also be detected in NPC cell lines (Cochet *et al.*, 1993). Previous studies have suggested that, unlike other EBV associated B-cell lymphomas where the virus is considered to be the initial factor of the tumour progression. In NPC the EBV infection only acts to facilitate or promote the tumour progression. EBV infection of healthy epithelial cells does not lead to NPC development, whereas EBV infection in epithelial cells unable to differentiate (due to other factors such as genetic or chemical carcinogens) provides the cells with growth and survival advantages, eventually leading to development of NPC (Shah & Young, 2009).

1.4. Next Generation Sequencing

It is becoming clear that the understanding of host and virus biology increasingly needs the determination of the host and pathogen genome, especially in virus tumours. Currently studies of viral and host interactions focus on single genes and their effect in infection or the disease progression. Little is known about the how the whole genomes interact with the host, and the interactions between multiple genes. Furthermore, current studies focus on the coding regions of the viral genome and little is known about the promoter and non-coding sequence of the viral genome, Lacking information of the whole viral genomes prevents further understanding of how the viruses truly interact with the host. Next Generation Sequencing (NGS) now allows further study of the whole viral genome; therefore it provides more information about variations in viral genomes which allows a fuller understanding of how virus function.

DNA sequencing has been a powerful and widely used tool in biological research. In 1977, Sanger et al. developed the dideoxynucleotide sequencing that led to modern capillary based methods used widely and almost exclusively in the last 30 years (Sanger *et al.*, 1977a). Sanger sequencing, (also known as dideoxy sequencing), uses both dideoxynucleotides (ddNTPs) and normal nucleotides (NTPs) in the process. ddNTP differs from NTP by containing a hydrogen group at the 3' carbon, instead of a hydroxy group (OH), this substitution prevents ddNTP from further incorporation with other NTP which results in the termination of the DNA chain. The resulting DNA fragments vary in length depending on the location of ddNTP incorporation. In capillary sequencing, different fluorescent dyes are used to label the ddNTP which

makes it possible to determine the nucleotide based on the termination position. The DNA sequence is determined by running the DNA fragments through size separating gel electrophoresis to separate by DNA length, and then using a laser to excite the fluorescent labels which provide the readout and DNA sequence. The Sanger sequence can achieve a DNA read length of to up 1000bp.

Following Sanger capillary sequencing of the Human Genome, many different approaches to sequencing have been developed in the last decade which aim to improve the rate and reduce the cost of sequencing. Improvements in other fields, such as computational data analysis, also enabled the development of the so called “Next Generation Sequencing” (NGS). Currently, NGS is used to refer to procedures developed using the cyclic-array sequencing technology, such as 454 sequencing (used by Roche Applied Sciences), Solexa technology (used by Illumina, referred to as Illumina sequencing), the SOLiD platform (used by Applied Biosystems), the Polonator and the HeliScope Single Molecular Sequencer technology (used by Helicos) (Shendure & Ji, 2008), with 454 and Illumina being the most commonly used platforms.

When compared with the traditional Sanger sequencing, there are several advantages associated with Next Generation Sequencing. Firstly, the array-based sequencing allows a higher degree of paralysation than the capillary based sequencing. Because the DNA fragments are immobilized to a planar surface during array-based sequencing, they can be manipulated with a single reagent volume, which lowers the cost of reagents. Secondly, array-based sequencing also allows deeper sequencing per base, which allows accurate base calling. Although 454 and Illumina are the most commonly used Next Generation Sequencing technologies, they rely on different PCR strategies to produce material for sequencing.

1.4.1. Illumina Sequencing

Prior to constructing the library for sequencing, DNA needs to be sheared to several hundred base pairs in length and ligated to adaptors. The Illumina sequencing relies on the bridging PCR method to generate amplicons. Both forward and reverse primers are tethered to a solid substrate at 5 prime ends by

a flexible linker, alternative cycles using Bst polymerase and denaturation with formamide after hybridisation results in clusters of PCR products, which consist of around 1000 clonal amplicons in each cluster (Figure 1.4.1.a). Several million clusters can be amplified in distinct locations with the eight independent “lanes” on a single Illumina flow cell, hence 8 independent libraries can be sequenced in parallel in the same run. After forming the clusters, the amplicons are denatured to single strands, allowing hybridisation of sequencing primers. Mixtures of four modified nucleotides along with the modified polymerase are added at each cycle of PCR. The modified nucleotides are also referred to as reversible terminators, due to the chemically cleavable moiety at the 3’ hydroxyl position, which prevents further binding and incorporation of other nucleotides (Turcatti *et al.*, 2008). An image of the four channels is taken after each single base extension, then both modified moieties are cleaved chemically prior to next cycle (figure 1.4.2.b) (Shendure & Ji, 2008). The base calling depends on the image showing the location of the florescent label. With Illumina sequencing, the original read pairs generated by Solexa technology generate short reads at around 36 bps. Every strand of DNA is sequenced twice, in the positive and negative senses, resulting in a read-pair. However, with recent models such as the Illumina HiSeq system, the read pairs generated are much longer to 100bps.

Figure 1.4.1. Illumina sequencing base on Solexa technology. a) The Illumina technology relies on bridge PCR (Adessi *et al.*, 2000; Fedurco *et al.*, 2006) (aka 'cluster PCR') to amplify clonal sequencing features. In brief, an *in vitro*–constructed adaptor-flanked shotgun library is PCR amplified, but both primers densely coat the surface of a solid substrate, attached at their 5' ends by a flexible linker. As a consequence, amplification products originating from any given member of the template library remain locally tethered near the point of origin. At the conclusion of the PCR, each clonal cluster contains 1,000 copies of a single member of the template library. Accurate measurement of the concentration of the template library is critical to maximize the cluster density while simultaneously avoiding overcrowding. (b) Each sequencing cycle includes the simultaneous addition of a mixture of four modified deoxynucleotide species, each bearing one of four fluorescent labels and a reversibly terminating moiety at the 3' hydroxyl position. A modified DNA polymerase drives synchronous extension of primed sequencing features. This is followed by imaging in four channels and then cleavage of both the fluorescent labels and the terminating moiety. Adapted from (Shendure & Ji, 2008).

1.4.2. 454 platform

The 454 platform developed by Roche was the first generation new sequencer to be commercialised. Any method that produces short-adaptor-flanked fragments can be used to generate the library. In contrast to the Illumina platform, clonal amplification is generated by emulsion PCR (Dressman *et al.*, 2003) with amplicons captured onto the surface of 28 μm beads. The primers are tethered to the beads at 5’ end in an oil and water emulsion PCR. Post amplicon caption, the untethered strands are removed from the beads. The beads covered with amplicons are hybridised to the sequencing primer (figure 1.4.2.a). Sequencing of 454 platform is performed by the pyrosequencing method (Ronaghi *et al.*, 1996), again using Bst polymerase. The beads are

deposited onto a microfabricated array of picoliter scale of wells, with diameters that only allow one bead per well. Smaller beads containing immoblized enzymes such as ATO sulfurylase and luciferase are the added to the array. These enzymes are essential for pyrosequencing. One side of the array acts as the flow cell for introducing or removing sequencing reagents, with the other side attached to a fibre-optic bundle for charge couple device (CCD) based signal detection. A single type of unlabelled NTPs is added to the array and, during the incorporation, pyrophosphate is released which acts *via* ATP sulfurylase and luciferase and results in burst of light detectable by CCD (figure 1.4.2.b). 454 technology allows generation of read-pairs which are 200-300 bps long.

Figure 1.4.2. Next Generation Sequencing of 454 platform. a)The 454 platform relies on emulsion PCR (Dressman *et al.*, 2003) to amplify clonal sequencing features. In brief, an *in vitro*–constructed adaptor-flanked shotgun library (shown as gold and turquoise adaptors flanking unique inserts) is PCR amplified (that is, multi-template PCR, not multiplex PCR, as only a single primer pair is used, corresponding to the gold and turquoise adaptors) in the context of a water-in-oil emulsion. One of the PCR primers is tethered to the surface (5'-attached) of micron-scale beads that are also included in the reaction. A low template concentration results in most bead-containing compartments having either zero or one template molecule present. In productive emulsion compartments (where both a bead and template molecule is present), PCR amplicons are captured to the surface of the bead. After breaking the emulsion, beads bearing amplification products can be selectively enriched. Each clonally amplified bead will bear on its surface PCR products corresponding to amplification of a single molecule from the template library. b) clonally amplified 28 μm beads generated by emulsion PCR serve as sequencing features and are randomly deposited to a microfabricated array of picoliter-scale wells. With pyrosequencing, each cycle consists of the introduction of a single nucleotide species, followed by addition of substrate (luciferin, adenosine 5'-phosphosulphate) to drive light production at wells where polymerase-driven incorporation of that nucleotide took place. This is followed by an apyrase wash to remove unincorporated nucleotide. Adapted from (Shendure & Ji, 2008)

1.4.3. Post Next Generation Sequencing analysis

Next Generation Sequencing is able to generate extremely large amounts of data. The bioinformatic analysis carried out to assemble the reads is crucial for successfully obtaining the sequence, and there are many different tools that can be used for this purpose. The choice of bioinformatics software depends on the purpose of the project. For whole genome virus sequencing, there are currently no reliable tools for carrying out *de novo* assembling, so more than one

software package is required to carry out the data processing. More details regarding analysis of data will be discussed in Chapter 4.

1.5. Aims of this Thesis

Despite the high prevalence of EBV and the multiple types of diseases that it can cause, little is known about the authentic cues that lead to the cellular induction of EBV replication and production of infectious virus in saliva. Since the virus establishes and maintains latency in non-proliferating memory B-cells, the signal to initiate the lytic cycle is unlikely to be from the virus (Hochberg *et al.*, 2004). Previously Laichalk and Thorley-Lawson (2005) have showed that terminal differentiation of B-cells into plasma cells can initiate the lytic cycle of EBV *in vivo*. If this occurs in the lymphoid tissue of the oropharynx, it would allow infectious virus to be shed into saliva (Yao *et al.*, 1989).

Previously, we and others have shown that XBP-1s is sufficient to initiate the lytic cycle of KSHV in PEL cell lines (Dalton-Griffin *et al.*, 2009; Sun & Thorley-Lawson, 2007; Wilson *et al.*, 2007). XBP-1s transactivates the promoter of KSHV immediate early protein RTA (ORF 50). Interestingly previous studies have also suggested that XBP-1s is able to bind to and transactivate the EBV BZLF1 promoter. However, XBP-1s may not be sufficient to reactivate EBV by itself since other studies have shown that protein kinase D (PKD) is also required (Bhende *et al.*, 2007). It has also been suggested that EBV and KSHV mutually inhibit each other. It is therefore interesting to further investigate the reactivation of EBV in mature B-cell backgrounds by XBP-1, with and without KSHV infection.

There have been many studies regarding the polymorphisms of EBV genes, either associated geographically, or with oncogenesis. However, little is known about the variations between EBV whole genomes. With limited numbers of whole genome sequence in Genbank, all either sequenced from a tumour sample, or after transforming B-cells; there is no EBV whole genome sequence from an EBV positive individuals with no malignancy. This thesis also investigates methods that have been developed to determine the EBV and/or KSHV full genome sequence from various different samples, *via* Next Generation Sequencing and further analysis the polymorphisms and variants from these EBV sequences.

Chapter 2

Materials and Methods

2.1. General Molecular Biology Techniques

2.1.1. Preparation of Competent bacteria

Two types of *Escherichia coli (E. coli)* were used in this study, XL-blue (Stratagene) and HB101. Either XL-blue or HB101 were streaked onto Luria-Bertani (LB)-agar plates with no antibiotic, and incubated at 37^{o}C for 16 hours. A single colony was picked and used to inoculate 5ml LB-broth starter culture and shaken at 37^{o}C for 16 hours. The starter culture was then added into 200ml LB-broth, and shaken at 30^{o}C until the absorbance at 595 nm was between 0.5 and 0.6. The culture was divided into 50ml falcon tubes and put on ice for 10 minutes, before being centrifuged at 3,500 rpm, for 10 minutes at 4^{o}C. The bacteria pellet was then re-suspended in 20ml cold TFB1 (30mM potassium acetate, 100mM rubidium chloride, 10mM calcium chloride, 50mM manganese chloride and 15% Glycerol) and incubated at 4^{o}C for 5 minutes. The bacteria was then centrifuged again at 3,500 rpm for 10 minutes, at 4^{o}C, re-suspended in 2ml TFB2 (10mM PIPES (pH 6.5), 75mM calcium chloride, 10mM rubidium chloride and 15% glycerol) and incubated at 4^{o}C for another 10 minutes. The bacterial suspension was quick frozen with liquid nitrogen in 100µl aliquots, and stored at -80^{o}C.

2.1.2. Introduction of plasmid DNA into *E. coli* by heat shock

1-200ng of plasmid was mixed with 100µl of competent XL-blue or HB101 *E. coli,* which was thawed on ice, and incubated on ice for 10 minutes. The mixture was then heat shocked at 42^{o}C for 45 seconds and immediately placed on ice for another 2 minutes. 200µl of LB-broth was added to the mixture, and incubated at 37^{o}C for 30 minutes with shaking at 250 rpm. 100µl of the culture was plated on LB-agar plates containing 50µg/ml ampicillin or kanamycin. Plates were incubated upside down at 37^{o}C for 16 hours.

2.1.3. Plasmid DNA midi-preps (Qiagen kit)

Midi-preps of plasmid DNA were produced from 50ml overnight cultures from either transformed XL-blue or HB101 *E. coli*, using the Plasmid Midi Kit (Qiagen) according to manufacturer's instructions.

2.1.4. Plasmid DNA mini-preps (Qiagen kit)

Mini-preps of plasmid DNA were produced from 3ml overnight cultures from either transformed XL-blue or HB101 *Escherichia coli*, using the Plasmid Mini Kit (Qiagen) according to manufacturer's instructions.

2.1.5. Molecular cloning

KOD hot start polymerase (Merck) with proof-reading activity was used for cloning. GoTaq polymerase with no proof-reading activity (Promega) was used for general PCR amplification. Restriction endonucleases were purchased from Promega. Calf intestinal alkaline phosphatase and T4 ligase were purchased from Promega. All were used in accordance with the manufacturer's specifications.

2.1.6. Cloning using the pGEM-T-EASY vector system (Promega)

The desired region was PCR amplified using 1 unit of KOD Hot start polymerase (Merck) in a 50µl reaction volume buffered with KOD Hot start polymerase 10x buffer. PCR conditions involved 20-100ng of template DNA and 10picomoles of each oligonucleotide primer. In addition, the reaction was supplemented with a final concentration of 0.2 mM dNTPs (Merck) and between 1.5 mM Mg++ (Merck). Cycling conditions were used in accordance with the manufacturer's instructions.

The desired PCR product was purified using a Qiagen gel extraction kit (Qiagen). Briefly 500µl of QC buffer was added directly to PCR reaction mix. 10µl of isopropanol was added and mixed well before all 650µl were applied to the column. The manufacturer's instructions for gel extraction were followed and

the final PCR result were eluted in 32µl. Poly-A tails were added to the purified PCR product prior to ligation to the pGEM-T-EASY vector system. The purified PCR product was incubated for 10 minutes with 1.5 units GoTaq polymerase and 1.6µl of 5nM dATPs in 8µl of GoTaq buffer with $MgCl_2$ and 28µl of PCR product at 72^{o}C.

The correct molecular weight product was separated by electrophoresis and visualised using 10mg/ml ethidium bromide (Sigma Aldrich). TAE buffered agarose gel with a percentage between 1-2% depending on the size of expected product was used. The product was subsequently excised and extracted using the MinElute gel extraction kit (Qiagen) in accordance with the manufacturer's instructions.

The following ligation reaction was used with the pGEM-T easy vector system (Promega). Negative controls were used by substituting dH_2O with PCR-amplified DNA and ligated in parallel.

5 µl 2x ligation buffer

1 µl T4 DNA ligase

0.5 µl pGEM-T-Easy linear DNA

3.5 µl MinElute extracted PCR product

Ligation reactions were transformed as described in chapter 2.1.2. Restriction enzyme digestion or PCR screening were used to confirm the positive colonies.

2.1.7. Sub-cloning

Primers were designed to contain restriction endonuclease sites to subclone the gene of interest from pGEM-T-Easy vector into appropriate expression construct or lentiviral vector. Vectors were digested overnight with appropriate restriction endonucleases in accordance with the manufacturer's instructions and treated with calf intestinal alkaline phosphatase (Promega) before being purified using the MinElute gel extraction system (Qiagen), according to the manufacturer's instructions. The pGEM-T Easy vector containing the gene of

interest was checked by digestion overnight with appropriate expression vector. Ligations were carried out using the following reaction.

5 µl 2x ligation buffer

1 µl T4 DNA ligase

1 µl linear alkaline phosphatase treated vecto DNA

3 µl MinElute extracted PCR product

2.2. Plasmid

2.2.1. pXBP1sG

The lentiviral vector construction of pXBP1sIG had been described previously (Wilson *et al.*, 2007). In brief, the pXBP1sG construct was developed by C. Tsantoulas, who generated it during an MSc project at the Kellam laboratory under the supervision of Dr. S. Wilson. The pXBP1sG construct encodes XBP-1s tagged with a c-terminal His-tag, cloned in the IRES-Em vector. XBP-1s tagged with a c-terminal His-tag cloned in pCDNA3 (GIBCO) was also supplied by S. Wilson.

2.2.2. pIG

The pIG plasmid is also referred to as pCSBX containing a IRES-Emerald encoding sequence, which is a kind gift from Y. Ikeda.

2.2.3. pPKDm-IG

The lentiviral vector for activated PKD, pPKDm-IG, was produced from the pHA.PKD.S738A/S742A plasmid kindly supplied by Alex Toker (Storz & Toker, 2003) (Addgene). The mutations at position 738 and 742 ensured the PKD was constitutively active once expressed. Briefly, BamHI and XhoI were used to clone the HA.PKD.S738A/S742A insert into the lentiviral vector pIG to produce the pPKDm-IG construct.

2.2.4. pZ-promoter-luciferase and pR-promoter-luciferase

The pZ-promoter-luciferase construct is a kind gift from Prof. Paul J. Farrell (Bryant & Farrell, 2002). The pR-promoter-luciferase construct is a kind gift of Prof. Shannon Kenney (Bhende *et al.*, 2007).

2.2.5. pBZLF1 and pBRLF1 expression construct

The pBZLF1 construct is a kind gift from Prof. Paul J. Farrell.(Amon *et al.*, 2004) and the pBRLF1 construct is a kind gift from Prof. Shannon Kenney (Bhende *et al.*, 2007; Bhende *et al.*, 2004).

2.2.6. Short hairpin RNA

The lentiviral vector expressing the short hairpin RNA (shRNA) of XBP-1 (Wilson *et al.*, 2007) and luciferase (Dalton-Griffin *et al.*, 2009) was generated previously. Briefly, the primers used for XBP-1 were (5′-CGA AAA AGA CTG CCA GAG ATC GAA AGA AGG CGA ACC TTC TCT CAA TCT CCG GCA GTC GGT GTT TCG TCC TTT CCA CAA GAT-3′ and T7 oligonucleotide 5′-TAA TAC GAC TCA CTA TAG GG-3′) and Luciferase (5’-CGA AAA AGG TGG CTC CCG CTG AAT TGG AAT CGA AAT CCC AAT CCA-3’ and 5’-GCG AGA GCC ACC GGT GTT TCG TCC TTT CCA CAA GAT ATA TAA AG-3’). Primers were used to PCR amplify a shRNA expression cassette using the plasmid pGEMU61Linker. The amplicon was subcloned into the pGEM-T-Easy vector system. Once the sequence was verified, it was cloned *via* EcoR1 restriction into an antibiotic resistance modified version of pCSPW namely the puromycin acetylase (Pac) encoding pHR-SIN-CSPW.

2.3. Cell Culture

All the cell culture work was carried out in class II biosafety cabinets, with sterile equipment. All the plates and containers used in this study were from Techno Plastic Products (TTP).

Table 2.1 The following B-cell lines were used in this study

Cell line	B-cell tumour	Viruses	Reference
JSC-1	Primary effusion lymphoma	KSHV and EBV	(Cannon *et al.*, 2000)
HBL-6	Primary effusion lymphoma	KSHV and EBV	(Gaidano *et al.*, 1996)
Cro-Ap6	Primary effusion lymphoma	KSHV	(Carbone *et al.*, 2000)
Cro6 clone 2	Primary effusion lymphoma	KSHV and EBV-GFP	(Trivedi *et al.*, 2004)
BC3	Primary effusion lymphoma	KSHV	(Arvanitakis *et al.*, 1996)
BC3 clone 6	Primary effusion lymphoma	KSHV and EBV-GFP	(Trivedi *et al.*, 2004)
Akata	Burkitt's lymphoma	EBV	(Takada, 1984; Takada *et al.*, 1991)
Raji	Burkitt's lymphoma	EBV	(Pulvertaft, 1964)
Mutu I	Burkitt's lymphoma	EBV	(Gregory *et al.*, 1990)
Daudi	Burkitt's lymphoma	EBV	(Klein *et al.*, 1967)
B95.8	LCL	EBV	(Miller *et al.*, 1972)
DG75	Burkitt's lymphoma	Negative	(Ben-Bassat *et al.*, 1977)

All the B-cell lines used in this study are grown in complete medium, Roswell Park Memorial Institute (RPMI) – 1640 medium (GIBCO, Invitrogen) with 10% foetal calf serum (FCS, BioSera) and 100units/ml penicillin plus 100µg/ml streptomycin (P/S, GIBCO, Invitrogen). With the exception of Cro6 clone 2 and BC3 clone 6, which are both grown in 20% FCS, with 1mg/ml and 2.5mg/ml G-418 sulphate (GIBCO, Invitrogen) respectively. All the cell lines were cultured at 37^{o}C and in 5% CO_2. The cells were split 1:10 or 1:20; depending on the density and the growth rate, twice a week.

Table 2.2 Cell lines of non-B-cell origin used in this study

Cell line	Description	Reference
HEK 293	293 cells containing SV40 T-antigen, a temperature sensitive gene	(DuBridge *et al.*, 1987)
HeLa	Eptheloid cervix carcinoma	(Masters, 2002)

HEK 293 cell line was cultured in Dulbecco's modified Eagle's medium (DMEM. GIBCO, Invitrogen) with 15% FCS and P/S, incubated at 37^{o}C and in 10% CO_2. HeLa cell line was grown using the same conditions as above except in 10% FCS and 5% CO_2. These cells were passaged three times a week and split at either 1:4 or 1:6.

2.3.1. Freezing cells

Cells were centrifuged at 325 x g (1200 rpm) for 5 minutes, and resuspended in cold FCS with 10% dimethyl sulphoxide (DMSO, Sigma), with a concentration at 5x10^{6} cells/ml. Cells were then aliquoted into cryovials (Nunc) and cooled to -80^{o}C in an isopropanol cryo-container (Nalgene) before being transferred and stored in liquid nitrogen.

2.3.2. Thawing cells

After removing the cells from liquid nitrogen, either 10ml of RPMI-1640 medium (for suspension cells) or DMEM medium (for adherent cells) with 10% foetal calf serum (FCS, BioSera) and 100 unit (U)/ml penicillin and 100µg/ml streptomycin (Invitrogen) was added to the cells. Cells were then pelleted at 325 x g for 5 minutes and resuspended in suitable volume of cultured medium, depending on the cell number.

2.4. Transient Transfection

2.4.1. Transient transfection of HEK 293 cells to make lentiviral vectors

Lentiviral vectors were produced as previously described by Besnier *et al*. (2002). HEK 293 cells were seeded such that 10cm dishes were confluent on the day of transfection. 10µl of FuGENE-6 (Roche) was added to 200µl of Opti-MEM (Invitrogen). 1µg of p8.91, 1µg of pMDG and 1.5µg of vector-genome encoding plasmid DNA were made up to 15 µl in Opti-MEM and added to the FuGENE-6 and Opti-MEM mixture. The mixture was incubated at room temperature for 30 minutes and then added to confluent HEK 293 cells in 8ml fresh DMEM medium (with 10% FCS), 100 units/ml penicillin and 100µg/ml streptomycin, drop wise and evenly distributed on the plate.

The plate was incubated at 37^{o}C in 10% CO_2 overnight and the medium was changed after 24 hours of incubation. The lentiviral vector containing supernatants were harvested at 48 and 72 hours post-transfection, and passed through a 0.45µM filter before being stored at -80^{o}C.

2.5. Lentivirus transduction and titration

2.5.1. Purification and concentration of lentivirus

The lentiviral vector containing medium was purified and the vectors were concentrated using polythylene glycol (PEG) 8000 (Sigma). After the lentiviral vector containing medium was passed through a 0.45µm filter, PEG 8000 and sodium chloride solution (5M, Sigma) were added to the solution with a final concentration of 5% and 0.15M respectively. The mixture was kept at 4^{o}C for a minimum of 24 hours to a maximum of 7 days. The virus was pelleted at 2000g at 4^{o}C for 10 minutes and the supernatant was discarded. The virus pellet was re-suspended in RPMI (10% FCS, 100 U/ml penicillin and 100µg/ml streptomycin) to 1/10th of the original volume and kept at -80^{o}C.

2.5.2. Lentiviral vector titration of infectious units by GFP expression in adherent cells

$1x10^5$ HEK 293 cells were seeded per well in a 6-well plate and cultured overnight at 37°C and in 10% CO_2, in 1ml of medium (DMEM containing 10% FCS, 100 U/ml penicillin and 100µg/ml streptomycin). The lentiviral-vector containing supernatant was diluted 1 in 3 in a serial dilution five times, starting with 250µl. 15µg/ml polybrene (Sigma) was also added to the cell-lentiviral vector mixture and the cells were incubated at 37°C and in 10% CO_2 for 48 hours. The percentage of GFP expressing cells was determined by flow cytometry relative to total number of cells.

2.5.3. Lentiviral vector titration of infectious units by GFP expression in suspension cells

$0.5x10^5$ JSC-1 cells was seeded per well in a 24-well plate and cultured at 37°C and in 5% CO_2 in 0.5ml of medium (RPMI containing 10% FCS, 100 U/ml penicillin and 100 µg/ml streptomycin). The lentiviral-vector containing supernatant was diluted 1 in 3 in a serial dilution five times, starting with 100µl. The supernatant was then mixed with RPMI medium to a total level of 500µl before adding to the JSC-1 cells. 15µg/ml polybrene (Sigma) was also added to the cell-lentiviral vector mixture, which was spun at 500x g for 1 hour at room temperature. The cells were then incubated at 37°C, in 5%CO_2 for 48 hours. The percentage of GFP expressing cells was determined by flow cytometry relative to total number of cells.

2.4.4. Flow cytometry

The FACS analysis was carried out on FACScan flow cytometers with Cellquest software (Becton Dickinson). A total of ten thousand events were collected. Infectious units were plotted against µl of supernatant. The absolute titre of the lentiviral vector was determined by extrapolating the infectious units per ml (IU/ml) using the straight-line equation of the above plot.

2.4.5. Preparation of live cells for flow cytometry

48 hours post infection; the adherent cells were washed once with PBS, and incubated with trypsin-EDTA (GIBCO, Invitrogen) at room temperature for 1 minute. The trypsin was quenched with DMEM supplemented with 10% FCS and P/S and transferred to capped FACS tubes (Falcon). The cells were pelleted and re-suspended in PBS with 3.3% paraformaldehyde. The suspension cells were washed with PBS once and re-suspended with PBS with 3.3% paraformaldehyde. Both adherent and suspension cells were kept on ice prior to analysis.

2.4.6 Antibiotic selection following lentiviral vector transduction

72 hours post transduction, the medium was removed and replaced with medium supplemented with 28μg/ml puromycine dihydrochloride (Sigma Aldrich) or 750μg/ml hygromycine B (GIBCO, Invitrogen), depending on the encoded resistance. The medium was removed and replaced with fresh medium supplemented with the appropriate antibiotic, every 2-3 days until non-transduced cells in parallel cultures were no longer viable. The antibiotic selected cells were then cultured normally in medium supplemented with the respective antibiotic.

2.6. Immunoblotting

48 hours after transduction (with no antibiotic selection), the samples were lysed with the mixture of DTT and sample buffer (1:3 ratio) (0.2M Tris-CL pH6.8, 5.2% SDS, 20% Glycerol and bromophenol blue), then sonicated and heated to 95^{o}C for 5 minutes. Protein concentration was normalised by blotting with housekeeping protein. Samples were resolved by 10% SDS poly-acrylamide gels and SDS-PAGE running buffer (10X in 50 litres: 50g SDS, 150g Tris, 720g glycine). The separated proteins were then transferred to a polyvinylidene fluoride membrane (PVDF) (GE healthcare) with either semi-dry or wet transfer apparatus using transfer buffer (25 mM Tris, 192 mM glycine, 20% methanol).

The membrane was blocked with 5% non-fat dried milk powder (NFDM) in TBS solution containing 0.1% Tween 20 (Sigma) (TBS-T) for 1 hour. The membrane

was then probed with primary antibody; BZLF1; BZ-1 diluted 1:2000 (Bryant & Farrell, 2002); Tubulin: anti-alpha Tubulin clone DM1A (Sigma), diluted 1:40,000; K-RTA: polyclonal, kind gift of Don Ganem (Lukac *et al.*, 1998), diluted 1:40,000; PKD: PKC μ (D-20): sc-935 (Santa Cruz), diluted 1:2000; BRLF1, anti-EBV transcription factor R (Argene) diluted 1:2000; BMRF1, EBV early antigen diffuse (Vector Labororatories) diluted 1:2000; XBP-1s, a kind gift from Giovanna Roncador (Maestre *et al.*, 2009), diluted 1:50), all in 1% NFDM in TBS-T solution at 4^{o}C overnight. We optimised the dilution of antibodies. The membrane was then incubated with appropriate HRP conjugated secondary antibodies (GE Healthcare) in 1% NFDM TBS-T solution for 1 hour at room temperature. The blots were washed 5 times for 5 minute periods in TBS-T before developing and visualising using ECL^{TM} Western blotting detection reagents or ECL Advanced Western blotting detection reagents (GE Healthcare).

2.7. 96-well plate Luciferase assay

HEK 293 cells or HeLa cells were plated at a density of 2x10^{4} cells per well in 96 well plates. Cells were transfected the next day with 20ng of the appropriate reporter and expression constructs using fuGNENE-6 (Roche). The ORF50 reporter plasmids were a kind gift from Erle Robertson (Robertson & Ambinder, 1997). The BZLF1 promoter and expression vectors have been described before (Bryant & Farrell, 2002) and the BRLF1 promoter and expression constructs were a kind gift from Shannon Kenney (Bhende *et al.*, 2007). The pXBP1sIG and pIG plasmid were described previously (Wilson *et al.*, 2007). Co-transfection of 2ng of Renilla expression vector was also used for monitoring transfection efficiency. After 48 hours, the relative light units and the expression of Renilla construct were determined using the luciferase Stop & Glo reagents according to the manufacturer's instructions. Data was read using a GloMaz96 microplate luminometer (Promega) with single injector. All assays were performed in triplicate.

2.7.1. Methylation of plasmids

Luciferase plasmids containing BZLF1 and BRLF1 were methylated using CpG methyltransferase (M. SssI) (New England BioLabs). 1µg DNA was mixed with 1µl of M. SssI (4 units/µl), 160 nM of S-adenosylmethionine (SAM) and buffer, according to the manufacturer's instructions. The mixture was then incubated at 37^{o}C for 1 hour. After incubation, 500µl of QC buffer from the gel extraction kit (Qiagen) was added to the mixture, with 100µl iso-propanol and following the manufacturer's instruction to extract the methylated DNA.

2.7.2. Methylation status of plasmids

The methylation of the plasmids was confirmed by performing restriction enzyme digestion with methylation sensitive enzymes. The Zp was digested with *BamH1* and *XhoI*; and the Rp was digested with *BglI* and *MlnI*.

2.7.3. Dual luciferase system

HEK 293-T cells or HeLa cells were plated at a density of 2x10^{4} cells per well in 96 well plates. Cells were transfected the next day with 20ng of the appropriate reporter and expression constructs using fuGNENE-6 (Roche). The ORF50 reporter plasmids were a kind gift from Erle Robertson (Robertson & Ambinder, 1997). The BZLF1 promoter and expression vectors have been described before (Bryant & Farrell, 2002) and the BRLF1 promoter and expression constructs were kind gift from Shannon Kenney (Bhende *et al.*, 2007). The pXBP1sIG and pIG plasmid were described previously (Wilson *et al.*, 2007). Co-transfection of 2ng of Renilla expression vector were also used for monitoring transfection efficiency. After 48 hours, the relative light units and the expression of Renilla construct were determined using the luciferase Stop & Glo reagents according to the manufacturer's instructions. Data was read using a GloMaz96 microplate luminometer (Promega) with single injector. All assays were performed in triplicate.

2.8. Detecting virion-associated genome copies using quantitative PCR

2.8.1. Inducing EBV and KSHV into lytic cycle

Cells were cultured at 2 x 10^5 /ml and treated with chemical stimuli: sodium butyrate (NaB) (Sigma) at 3mM; valporic acid (VPA) (Merck) at 1.5mM; and 12-*O*-tetradecanoyl phorbol 13-acetate (TPA) (Sigma) at 20ng/ml. Goat anti-human IgG and anti-human Ig-M (Sigma) were both used at 5ug/ml. All but NaB were incubated overnight at 37^{o}C, NaB was washed off after 1 hour incubation.

2.8.2. DNA extraction

The DNA was extracted from 2-10 x 10^5 cells using QIAmp DNA extraction kit (Qiagen) according to the manufacturer's protocol. 200μl of saliva samples or concentrated virus supernatant were used to extract viral DNA using QIAmp blood DNA extraction kit (Qiagen) according to the manufacturer's protocol. 10ug of salmon sperm DNA (ssDNA) was added to the sample prior the extraction as a carrier for some samples. The total elution volume was 32μl.

2.8.3. Quantitative PCR (q-PCR)

Viral load was measured by quantitative PCR (q-PCR) using SensiMix dU kit (Bioline) with 5mM $MgCl_2$. The primers were used at 20pmol (EBV forward primer 5'-GGC CAG AGG TAA GTG GAC TTT AAT-3', EBV reverse primer 5'-GGG GAC CCT GAG ACG GG-3'); the probe was used at 10pmol (EBV probe: 5' FAM-TCT TCT CAA AGG CCA CCG CTT TCA AGT C-TAMRA 3'). The primers were a kind gift from Dr. Paul Grant and target the EBNALP gene. The reactions were run on an ABI 7000 thermocycler with an initial 10 minutes incubation at 37°C for UNG, then followed by another 10 minutes incubation at 95°C, followed by 40 cycles at 95°C for 15 seconds and 60°C for 1 minute. Every sample was run in duplicate and the viral copy number per μl was calculated using a standard curve that was run alongside the samples. The standard curve of EBV was calculated by using samples with a known copy number of EBV and its relationship with Ct numbers.

2.9. RT-PCR

2.9.1. RNA extraction and DNase treatment

RNA was extracted using the Trizol reagent (GIBCO, Invitrogen). $1x10^6$ cells were lysed per ml of Trizol reagent in accordance with the manufacturer's instructions. The cell lysate was stored at -80^oC until used for RNA extraction. The frozen cell lysates were thawed on ice prior to repeated chloroform extraction. The RNA was extracted and purified using RNeasy mini extraction kit (Qiagen) after an on-column DNase (Promega) digestion. Both were carried out in accordance with the manufacturer's instructions. RNA was quantified using ND-1000 spectrophotometer (Nanodrop Technologies)

2.9.2. cDNA synthesis

The reverse transcription reaction was carried out using Omniscript Reverse Transcriptase (Qiagen), in accordance with the manufacturer's instructions. The starting quantity of total RNA was between 1-2µg.

2.9.3. Detecting messenger RNA

2.9.3.1. XBP-1 PCR, *Pst* I digestion and resolution

PCR amplification was performed across the atypical sliced junction of XBP-1 by using primer XBP-1 short fwd (5'-CCT TGT SGT TGA GAA CCA GG-3') and XBP-1 short rev (5'-CAG AAT GCC CAA CAG GAT ATC-3') at 20pmol. Total PCR reaction volume was 50µl with 5 units of GoTaq polymerase and 5x GoTaq polymerase flexi buffer (Promega). The reaction was hotstarted at 95^oC for 2 minutes, followed by 35 cycles at 95^oC for 30 seconds, 58^oC for 30 seconds and 72^oC for 1 minute.

Following the PCR amplification, the 50µl reaction was halved to 2 x 25µl samples. One was added to 1µl dH_2O and the other to 20 units of *PstI* restriction enzymes. Both were incubated at 37^oC for 1 hour. 3% agarose gel with TAE buffer was used to resolve the PCR products, at 5 volts/cm at room temperature.

2.9.3.2. q-PCR to detect messenger RNA level

Copies of messenger RNA (mRNA) were measured by q-PCR after reverse transcription (q-RT-PCR). The experiments were carried out using SyBr Green PCR kit (Qiagen) according to the manufacturer's protocol. All the primers used are listed below in table 2.3 and were used at 5mM. The reactions were run on an ABI 7000 thermocycler at 95^{o}C for 15 minutes, followed by 40 cycles of 95^{o}C for 15 seconds, and 60^{o}C for 60 seconds. Known copies of the viral DNA were run alongside each experiment to create a standard curve. This was correlated with the Ct number, from which the viral copy number was calculated. All the samples were run in duplicate and the copy number of mRNA was normalised to L32, a house-keeping gene.

Table 2.3. list of primers used in q-RT-PCT

Name of primers	Sequence (5'- 3')	Reference
Z-F (BZLF1 qPCR forward primer)	CTATCAGGACCTGGGAGGGC	(Schelcher *et al.*, 2005)
Z-R (BZLF1 qPCR reverse primer)	CACAGCACACAAGGCAAAGG	(Schelcher *et al.*, 2005)
R-F (BRLF1 qPCR forward primer)	AATTTACAGCCGGGAGTGTG	(Chia *et al.*, 2008)
R-R (BRLF1 qPCR reverse primer)	AGCCCGTCTTCTTACCCTGT	(Chia *et al.*, 2008)
M-F (BMLF1 qPCR forward primer)	GGACCTGCCGTTGGATCTTA	(Schelcher *et al.*, 2005)
M-R (BMLF1 qPCR reverse primer)	TCTTCGGAGGCGTGGTTAAA	(Schelcher *et al.*, 2005)
ORF50-F (ORF50 qPCR forward primer)	TTGGTGCGCTATGTGGTCTG	(Caselli *et al.*, 2005)
ORF50-R (ORF50 qPCR reverse primer)	GGAAGGTAGACCGGTTGGAA	(Caselli *et al.*, 2005)
L32-F (house-keeping qPCR forward primer)	CAACATTGGTTATGCAAGCAACA	(Schelcher *et al.*, 2005)
L32-R (house-keeping qPCR reverse primer)	TGACGTTGTGGACCAGGAACT	(Schelcher *et al.*, 2005)

2.10. Sample preparation for deep sequencing

2.10.1. Gardella gel

The Gardella gel consists of 2 parts, the running gel (0.75% agarose gel) and the lysis gel (0.8% agarose gel, 2% SDS and 1mg/ml protease K), which contains the wells for sample loading. 2 x 10^6 cells were re-suspended in 50µl loading buffer (15% Ficoll in TBE buffer, 40µg/ml RNase A and 0.01% bromophenol blue). 25µl samples were loaded per well and the gel was run in TBE buffer at 4^{o}C, 40 volts for 2 hours and 160 volts for 15 hours. The gel was

then soaked with TBE buffer containing 0.5ng/ml ethidium bromide (EtBr) for 2 hours, before being exposed under UV light to determine the location of episomal DNA. The running of the gel was repeated and instead of soaking the gel with EtBr TBE buffer, the running gel was cut into 0.5cm length, and soaked with 200µl dH2O to elute the episomal DNA.

2.10.2. Phi29 PCR

Phi29 DNA polymerase (GE healthcare) and EBV specific primers (designed by Dr. Dan Frampton), with no cross-link with either human or KSHV genome (table 2.4) were used. 1µl DNA (approximate 10 ng) were used, with a 4µl primer mix (see table X, 5 µM/ml) and 5µl annealing buffer (80mM Tris-HCL, pH 8.0 and 20mM $MgCl_2$). The mixture was incubated at 95°C for 3 minutes then cooled at 4°C for 1 minute. After incubation, 1µl Phi29 DNA polymerase (Fermentas), 2µl 10X buffer, 1µl pyrophosphtase (0.025 units/ µl), 1µl BSA and 5µl of dNTP (25 mM) were added. The final mixture was incubated at 30^{o}C for 12 hours, 60^{o}C for 10 minutes and 4^{o}C for 5 minutes. The EBV copy number was determined by q-PCR.

Table 2.4. Primers used in Phi29 PCR

Primers and position	Sequence (5'- 3')
EBV-142	GGTGGCGGTCACAACGGTACTAA*C*T
EBV-2158	CATCTCCACGCGAAGCACGTACG*T*G
EBV-47323	GAGCGTAACTCGCCTCGTCTTCG*G*T
EBV-62291	GTCACCGTATACGCGGAGACCAA*C*C
EBV-9178	CTGTGTAGCTACCGATAAGCGGA*C*C
EBV-54317	CATCGTGAGCGAGTACGTGGATC*G*G
EBV-58831	CACATATACCGCTCGGCGGAGAT*C*T
EBV-70342	GTCATATCGGATGCCTCACGCGA*A*G
EBV-87302	CAGCTCATCGTCCGAGGACTCTA*G*C
EBV-92587	CAACGACGCATGTTCCATCGCGT*G*C
EBV-100664	CTCGCAGTTCGCCGACCAAGACA*T*G
EBV-124903	CCTGAGTCGTATGACGGATGGCA*A*G
EBV-125359	GATAACGCCGAGTAGAAGCCGAC*G*C
EBV-134002	GGCTATTGACTGGCGCGCTTCCT*T*G
EBV-153457	GATCCACGACCGCATCCAGTACG*T*C
EBV-155710	GAAGGTCACGCGCCGTTCCATTA*T*G
EBV-159157	CATCGAGCCGAGCGACTGGATAG*A*G
EBV-164663	GACACGCACGTACGGCTCATCAT*C*A

2.10.3. Genomiphi (Whole Genome Amplification)

The quantity of DNA samples was increased using Illustra GenomiPhi V2 DNA Amplification Kit (GE healthcare) in accordance to the manufacturer's protocol and purified using Zymoclean Genomic DNA Clean & Concentrator^{TM} (Zymo Research Corp.).

2.10.4. SureSelect Target Enrichment

Overlapping 120-mer RNA baits (5x coverage) spanning the length of the positive strand of the reference genomes were designed using Agilent eArray software. The specificity of all baits was verified by BLASTn searches against

the Human Genomc + Transcript database. The bait library consists of two sets of baits from EBV type 1 and EBV type 2 reference genomes each.

DNA preparations of 3µg were sheared for 6x60 seconds using a Covaris E210 (duty cycle 10%, intensity 5 and 200 cycles per burst using frequency sweeping). End repair, non-templated addition of 3′-A, adaptor ligation, hybridization, enrichment PCR and all post-reaction cleanup steps were performed according to the SureSelect Illumina Paired-End Sequencing Library protocol (Version 1.0) observing all recommended quality control steps. However, the Sanger institute multiplex 10bp tag was used at the end of the process.

2.11. Sequencing

2.11.1. DNA Sequencing

DNA Sanger capillary sequencing was used to confirm PCR and cloning product, and was carried out by either Cogenics (Lark) or the University College London sequencing facility. The results were analysed by using DNA dynamo program.

Table 2.5. Primers used to confirm the majority consensus sequence generated by Next Generation Sequencing.

Primers	Sequence (5'- 3')
BRLF1-1 F	ATGAGGCCTAAAAAGGATGG
BRLF1-1 R	CTAAAATAAGCTGGTGTCAAAAATAG
BRLF1-2 F	GCTACTCGGAGTCCGGACAG
BRLF1-2 R	GTAAAAGCATGAAACTGTCCGG
BMRF1 F	ATGGAAACCACTCAGACTCTC
BMRF1 R	TTAAATGAGGGGGTTAAAGGCC
BZLF1 F	ATGATGGACCCAAACTCGAC
BZLF1 R	TTAGAAATTTAAGAGATCCTCGT
BDLF3.5 F	TCAATCGGCCTTGGTCTGAC
BDLF3.5 R	ATGTCTGCCCCCGGATGC
EBNA2 F	TCTGCTATGCGAATGCTTTB
EBNA2 R	AGAGGGTGCATTGATTGGTC

2.11.2. Second generation sequencing (deep sequencing)

Second generation sequencing was carried out with an Illumina machine, at the Welcome Trust Sanger Institute, Hinxton, Cambridge, UK. All the cell line samples were sequenced using the Illumina GAIIx system, and the PTLD and saliva samples were sequenced using the Illumina HiSeq system, six samples per lane.

2.11.3. Computational Analysis of Second Generation sequencing Data

For each dataset, all duplications were removed from the read-pairs, then subject to quality control using QUASR v7 pipeline (http://sourceforge.net/projects/quasr/). Firstly the 3' end of the reads were trimmed to ensure that the median Phred quality score of the last 15 bases exceeded 30, and subsequently to remove the read-paired if either the read of the median Phred quality score was below 30 or less than 50bp in length. The remaining read-pairs were mapped back to the according reference genome using the Burrow-Wheeler Alinger (maximum insert 50 bases, maximum distance between paired ends 500) (Li & Durbin, 2009), generating SAM files containing all mapped and unmapped reads. The SAM files were processed using SAMTools (Li *et al.*, 2009) to produce pileup files for consensus sequence generation and SNP calling using VarScan v.2.2.7 (--min coverage 3, --min reads2 3, --p value 5e-02) (Koboldt *et al.*, 2009). The unmapped read paired were extracted from SAM files and BLASTn searches used to determine the proportion mapping to the reference genome.

2.11.4. Phylogenetic tree

Both type 1 and type 2 EBV majority consensus sequences and EBV GD1, GD2 (HQ020558) sequences were aligned against EBV type 1 reference sequence (NC_007605.1) using Base-By-Base software and Gendoc. Repeated regions of the sequences were removed according to the location on EBV type 1 reference genome. The coding region was extracted according to the location of EBV type 1 reference genome. The heterogeniety site model was determined using JModelTest with default setting, and Baysian skyline coalescent model (Drummond *et al.*, 2005) was used to construct two trees with forced roots,

relaxed longitudinal and relaxed exponential clock models, both sampling lengths ranging between 20-50 millions. Both relaxed clock models were run on default setting (initial value 1, lower value 0, upper value 1). The outcome of the relaxed longitudinal tree was used to determine if strict clock tree was more suitable. A second method using a distance based method: Neighbour joining was used to generate an unrooted tree, using MEGA software in default setting. This method was a bottom-up clustering method. The possibility of recombination between the sequences was examined using Recombination Detection Programe v3.3.1 (RDP3) software (Martin *et al.*, 2010), including RDP, BOOTSCAN, Chimaera, MaxChi and SiScan methods (Gibbs *et al.*, 2000; Martin & Rybicki, 2000; Martin *et al.*, 2005; Posada & Crandall, 2001; Smith, 1992).

Chapter 3

Results: XBP-1 expression and the effect on EBV reactivation in B-cell Lymphomas

3.1. Introduction

EBV lytic reactivation is mediated by two immediate early viral proteins: BZLF1 (ZTA) and BRLF1 (RTA) (Amon *et al.*, 2004; Chang & Liu, 2000; Yuan *et al.*, 2006). The Z promoter is activated after BCR cross-linking, but this is reported not to activate the R promoter (Rp) directly (Amon et al., 2004), indicating that BZLF1 is the first protein that is expressed after reactivation of EBV from latency. BZLF1 expression is also sufficient to initiate the entire EBV lytic cycle (Countryman & Miller, 1985; Rooney *et al.*, 1989b). KSHV-RTA is the virus immediate early replication and transcription activator protein, and has a similar role to EBV BZLF1. KSHV-RTA is also sufficient and necessary to activate the virus lytic cycle and transactivate its own promoter, which leads to the KSHV lytic gene expression cascade (Lukac et al., 1998).

XBP-1 spliced (XBP-1s) is a B-cell terminal differentiation transcription factor and is essential to transform B-cells into plasma cells at the final stage of B-cell development. EBV lytic replication is induced in plasma cells (Laichalk & Thorley-Lawson, 2005), however, in the environment where XBP-1s is expressed abundantly, such as multiple myeloma cell lines, infection by EBV does not result in lytic reactivation (Anastasiadou et al., 2009).

Previous work carried out by us and others has shown that XBP-1s is sufficient to initiate the lytic cycle of KSHV in PEL cells (Dalton-Griffin *et al.*, 2009; Sun & Thorley-Lawson, 2007; Wilson *et al.*, 2007). XBP-1s transactivates the KSHV immediate early promoter ORF50, leading to reactivation from latency. Here we investigate further the role of XBP-1s in reactivation of EBV in a mature B-cell background, namely PEL and BL cell lines. Other studies have shown that murine XBP-1s can induce BZLF1 protein expression *in vivo,* by transactivating the BZLF1 promoter (Sun & Thorley-Lawson, 2007). However it has also been shown that the reactivation of EBV requires not only XBP-1s expression, but also the expression of protein kinase D (PKD) (Bhende et al., 2007). In this

chapter, we investigate the interactions between human XBP-1s and EBV lytic reactivation.

3.2. Results

3.2.1. Over expressing XBP-1s in PEL and BL cell lines

Over-expression of XBP-1s protein was used to examine its effect on the reactivation of latent gamma-herpes virus in PEL and BL cell lines. Transduction using lentiviruses containing XBP-1s was used for over-expression. This lentiviral vector contains XBP-1s (pXBP1s-GFP) and was constructed by a previous group member in the laboratory and has shown successful XBP-1s over-expression in PEL cells, i.e. JSC-1 cells (Wilson et al., 2007). This lentiviral vector expresses XBP-1s from the SFFV promoter with an internal ribosome entry site (IRES) immediately after XBP-1s and before the GFP gene (Wilson *et al.*, 2007). FACS analysis was used to determine the rate of transduction of PELs and BLs

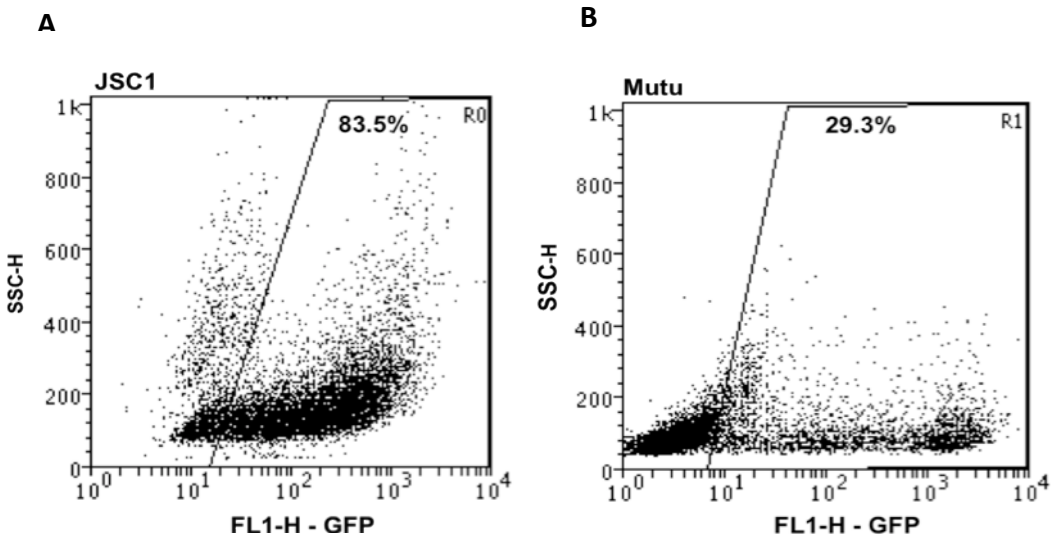


Figure 3.1. FACS analysis of lentiviral transduction, showing the percentage of A) JSC1 (PEL cell line) and B) Mutu I (BL cell line) that is GFP positive, after transduction of XBP-1s GFP lentivirus.

3.2.2. Determining the expression of BZLF1 and KSHV-RTA

In order to determine whether XBP-1s induced the lytic cycle of EBV, the expression of BZLF1 was used as an indicator of initiation of the EBV lytic cycle. Several different methods have been used to determine BZLF1 and K-RTA protein expression in both PEL and BL cell lines, such as FACS analysis. However, intra-cellular staining for the immediate early proteins with various

antibodies did not give clear results of the protein expression, either by FACS or confocal microscopy. Western blot analysis was used to show the expression of the viral proteins from the chosen cell lines. Daudi, a EBV positive BL cell line was been treated with TPA to reactivate EBV and therefore expressed BZLF1. The BL cell line DG75, both EBV and KSHV negative, was used as a negative control for the Western blot analysis. A lentiviral construct containing KSHV ORF50 was transfected into HEK-293 cells to express the K-RTA protein. Figure 2 shows clear BZLF1 and KSHV RTA protein expression from the Western blot.

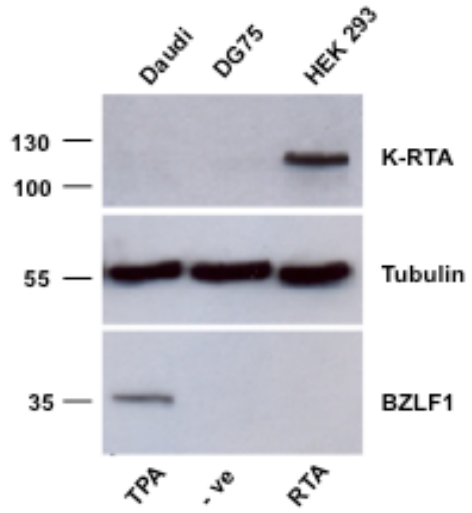


Figure 3.2. Using Western blot analysis to determine reactivation of EBV and KSHV. BZLF1 and K-RTA (RTA in lane 3) are immediate early protein of EBV and KSHV respectively. The expression of these two proteins can be detected using Western blot analysis, which indicate induction of lytic cycle.

3.2.3. XBP-1s over-expression in PEL cell lines results in KSHV RTA expression but not EBV BZLF1 expression

Over expressing XBP-1s in PEL cell lines (single or dually infected) induced K-RTA expression, was previously described by Wilson et al. (2007) and is also shown in figure 3.3. XBP-1 is unspliced and inactivated in the PEL cell lines, which can be induced into splicing by ER stress. Wilson et al (2007) have shown that in both JSC1 and BCP1 PEL cell lines XBP-1s is able to induce K-RTA expression by transactivating the ORF50 promoter. K-RTA expression alone is sufficient to induce KSHV into lytic cycle. In dually infected PELs, the effect of XBP-1s expression on EBV BZLF1 protein expression was not known.

Here, we over-expressed XBP-1s exogenously in PELs both singly and dually infected, such as Cro-Ap6 (CRO6) and JSC-1 cells respectively. TPA, a phorbol ester, is known to be able to induce lytic reactivation of both KSHV and EBV and is used as a positive control (figure 3.3.A lane 4, and B, C lane 2) to ensure that both viruses can be induced into lytic cycle. K-RTA, in both singly and dually infected PELs, was expressed after XBP-1s expression but not in the vector control (Figure 3.3.A and B, lane 2 and 3). However, in JSC1 cells, no protein expression of BZLF1 was detected, despite the ability of TPA to induce the expression of both K-RTA and BZLF1 simultaneously (figure 3.3.B). As described in chapter 3.2.2, Daudi treated with TPA, DG75, and HEK 293 cells transiently transfected with RTA were used as a Western blot control.

Since little is known about the EBV genome in dually infected PELs, it is possible that the EBV is somehow defective and non-responsive to XBP-1s. Therefore, we repeated the experiment on BC3 clone 6 (BC3 cl6) and CRO6 cone 2 (CRO6 cl2). These two cell lines were created by superinfecting KHSV positive only PEL cell lines (BC3 and CRO6) with replication-competent EBV-GFP virus (Xu et al., 2007) and selected with G418. Western blot analysis showed identical results for K-RTA and BZLF1 expression as JSC-1 cells (figure 3.3.B, C), K-RTA was induced after XBP-1s expression but not BZLF1. As before, TPA treatment confirmed that both KSHV and EBV can be induced into lytic cycle in these dually infected PELs. In these experiments, 60-80% of PELs in culture were transduced with XBP-1s expressing lentivirus, and the protein expression of XBP-1s is also confirmed by Western blot analysis (figure 3.3.A, B).

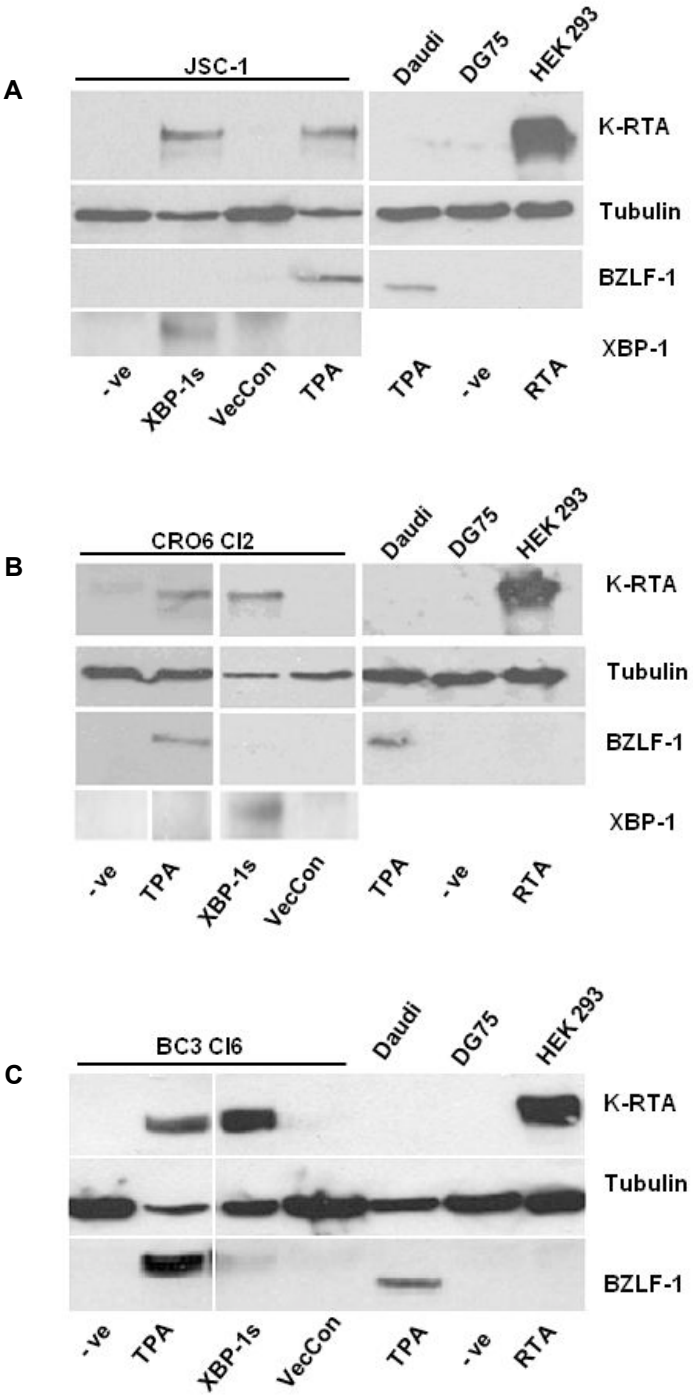


Figure 3.3. XBP-1s does not activate EBV lytic gene expression in PEL cell lines. (A) JSC-1, a KSHV and EBV double positive PEL cell line, expressed K-RTA and BZLF1 after treatment with TPA, but only K-RTA was expressed following XBP-1s transduction. VecCon indicates the vector control. B) CRO6 cl2, and C) BC3 cl6 are EBV superinfected KSHV positive PEL cell lines and showed a similar pattern with XBP-1s transduction only resulting in K-RTA expression. TPA was able to induce both K-RTA and BZLF1. The Daudi cell line treated with TPA and KSHV RTA (K-RTA) transfected HEK 293 cells acted as positive control for the BZLF1 and K-RTA antibody staining respectively. XBP-1s expression was detected using an XBP-1s specific antibody.

3.2.4. Over-expression of XBP-1s in BL cells does not induce EBV BZLF1 expression

It is possible that the presence of KSHV interferes with the ability of EBV to enter the lytic cycle and a previous study has suggested possible antagonistic interaction between KSHV and EBV (Miller et al., 1997). To investigate the effect of expression of XBP-1s on EBV lytic reactivation, without the presence of KSHV, we decided to over-express XBP-1s in BL cell lines which are singly infected by EBV. The transduction rate of BL cell lines was between 10-25%, Despite this not being as high as in PEL cell lines, inducing EBV lytic cycle should still be observable. Previous reports have suggested that XBP-1s binds to the Zp1 site on the Zp, and therefore transactivate the promoter, resulting in expression of BZLF1 and initiation of the lytic cycle (Sun & Thorley-Lawson, 2007). However, over expressing XBP-1s in two different BL cell lines, Mutu I and Akata, failed to induce BZLF1 expression (Figure 3.4.A and B). The ability of EBV to enter lytic cycle in these cells was investigated by using BCR crossing linking. Under these conditions BZLF1 expression can be detected by Western blot analysis. These results suggested that XBP-1s does not transactivate the BZLF1 promoter in B-cell lymphomas, regardless of the presence of KSHV.

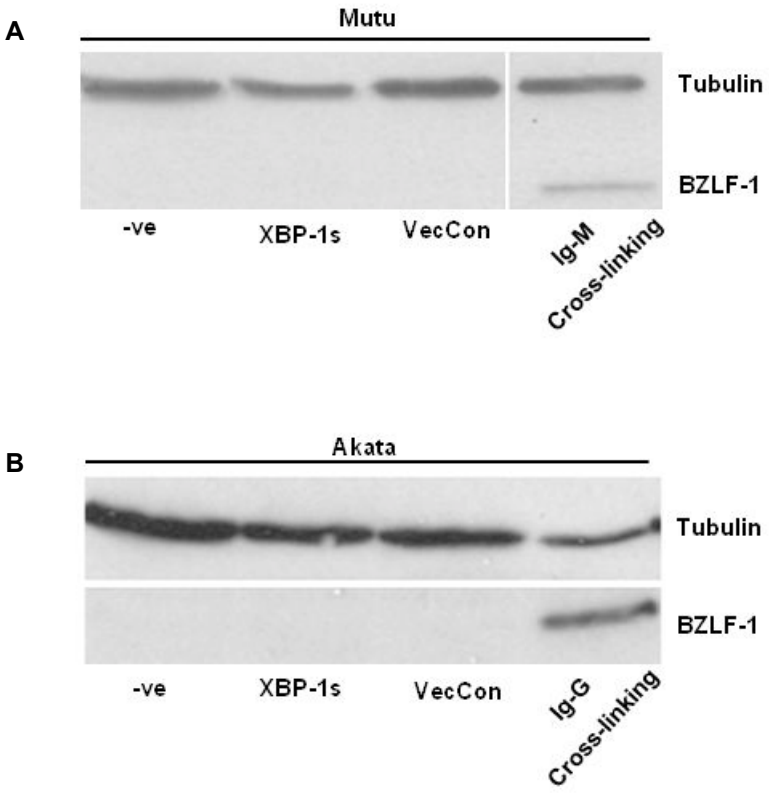


Figure 3.4. Burkitt's Lymphoma cell lines do not express BZLF1 after transduction with XBP-1s. A) Mutu I and B) Akata, both EBV positive Burkitt's lymphoma cell lines, were transduced with XBP-1s and vector control (VecCon) but BZLF1 expression was not detected. B-cell surface cross-linking (BCR) with human Ig is used as positive control to induce BZLF1 expression. A) Mutu I, with Ig-M. B) Akata, with Ig-G.

3.2.5. Over-expression of XBP-1s in BL or PEL cells does not induce EBV BRLF1 and BMRF1 expression

BZLF1 is the first protein that is expressed for the initiation of the lytic cycle of EBV. Nevertheless, there are two immediate early (IE) proteins for EBV reactivation. BZLF1 protein expression transactivates the BRLF1 promoter, leading to the expression of BRLF1, the second IE protein. Together, these two proteins can induce the full EBV lytic cycle cascade. We therefore examined the expression of BRLF1 after transducing XBP-1s in both BL (Akata) (Figure 3.5.A) and PEL (JSC-1) (Figure 3.5.B) cell lines. In both cell lines, both BZLF1 and BRLF1 expression can be detected after either treating the cells with BCR-crosslinking (Akata) or TPA (JSC1). However, after transducing both cell lines with XBP-1s, neither BZLF1 nor BRLF1 expression can be detected.

We also examined the expression of BMRF1, an EBV early protein also known as diffused early antigen (EA-D) (Bayliss & Wolf, 1981). BMRF1 transcription is activated by BZLF1 (Kenney et al., 1992). The expression of BMRF1 was examined using Western blot analysis, however, in both BL and PEL cell lines, over-expressing XBP-1s did not induce BMRF1 expression. Therefore, over-expression of XBP-1s in both PEL and BL cell lines does not induce the expression of either the IE proteins: BZLF1 and BRLF1, which are essential for inducing EBV lytic cycle; nor the expression early protein BMLF1. This indicates that XBP-1s over-expression does not induce EBV lytic cycle in PELs and BLs.

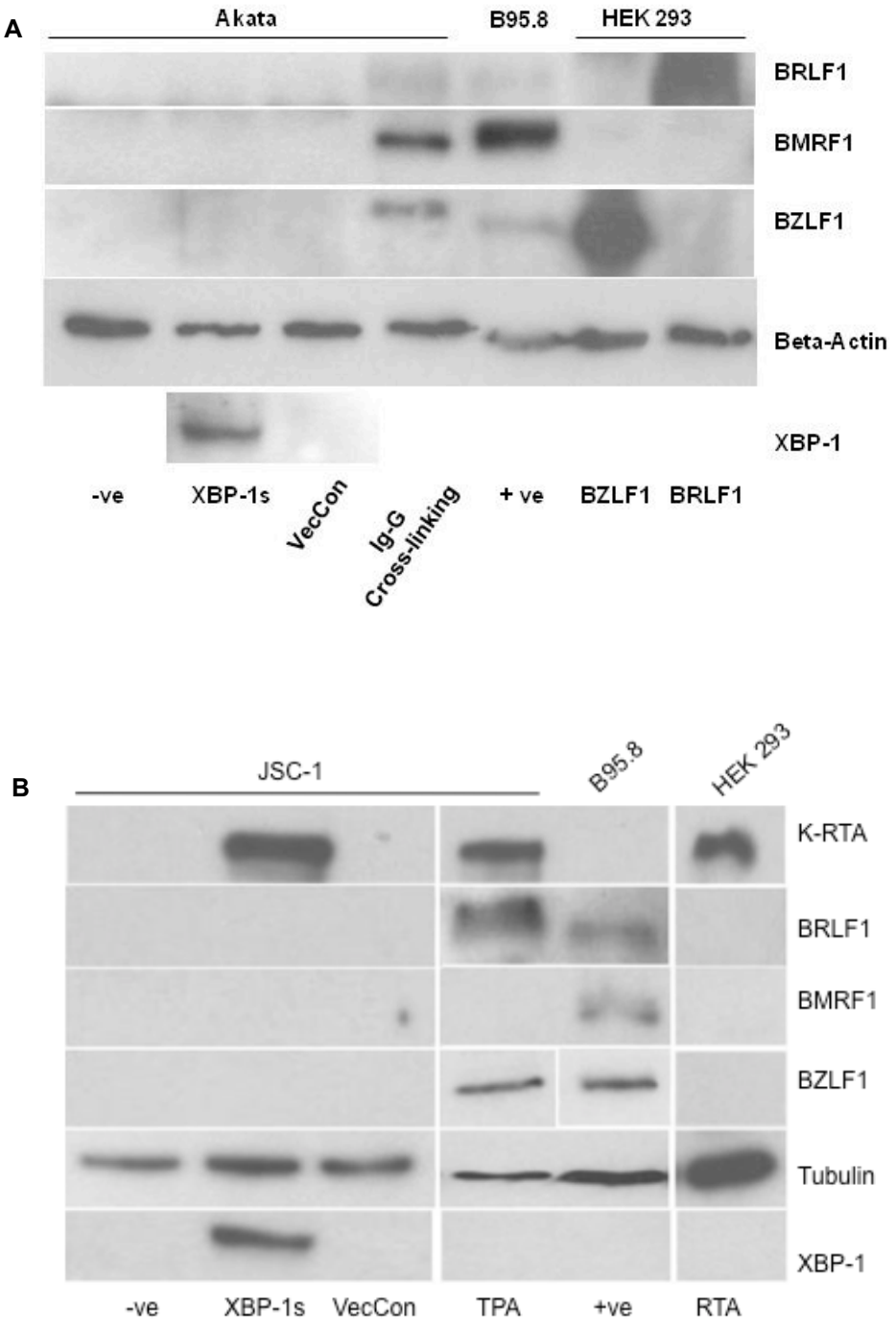


Figure 3.5. Expression of XBP-1s does not induce BZLF1 and BRLF1 protein expression in either BL or PEL cell lines. A) Akata and B) JSC-1 cell lines were transduced with XBP-1s and vector control with MOI of 5 and 2 (MOI measured in JSC-1 cells) respectively. Neither BZLF1, BRLF1 nor BMRF1 expression were detected in either cell lines. BCR cross-linking was used as positive control for Akata to induce EBV lytic gene expression and TPA was used for JSC-1 cell lines to induce KSHV and EBV lytic gene expression. Both positive controls show strong BZLF1 expression, weaker BRLF1 expression, and but only (A) Akata with BCR cross-linking expresses BMRF1. B95.8 cell line was used as positive control for EBV immediate early and early protein expression.

3.2.6. EBV lytic cycle promoters do not respond to XBP-1s.

Previously we showed that XBP-1s can transactivate the K-RTA promoter (Wilson et al., 2007) which induces the full KSHV lytic cycle. Despite showing that over-expression of XBP-1s in PEL and BL cell lines does not induce BZLF1 and BRLF1 protein production, we examined the effect at the promoter level of XBP-1s expression. A construct containing the BZLF1 promoter region fused with a luciferase reporter gene (McDonald et al., 2010) (a kind gift of Prof. Paul J. Farrell) was used to determine the response of XBP-1s in HEK 293 cells. We failed to see any strong XBP-1s activity on the BZLF1 promoter (Zp) (figure 3.6.A), especially compared to the strong response to BZLF1 protein. This weak response was also statistically insignificant ($P > 0.05$, two tailed T-test). Also, the BZLF1 promoter only responded to the positive control (BZLF1 protein expression) when the construct was methylated (black bar). BZLF1 protein expression failed to transactivate the Zp while it was unmethylated (data not shown). During thesse stages of EBV life cycles, the DNA is heavily methylated during latency but unmethylated at the lytic cycle (Robertson & Ambinder, 1997). BZLF1 is the first IE protein that is expressed in the lytic cycle and as EBV DNA remains methylated during latency, prior BZLF1 expression. Therefore, it is not surprising that Zp is only responsive to BZLF1 only when methylated.

These experiments were repeated using the BRLF1 promoter (Rp) (Bhende et al., 2007). Again, the experiments were carried out with BRLF1 promoters either methylated (black bar) or unmethylated (white bar). Both forms of the promoter were able to be transactivated by BRLF1, unlike the BZLF1 promoter. We were also able to detect, but to lesser level, a response to XBP-1s (figure 3.6.B). Therefore BRLF1 promoter activity is independent of the promoter methylation state and XBP-1s was able to weakly transactivate BRLF1 promoter in HEK 293 cells. All results were statistically confirmed by using two-tailed T-test, to confirm the increase in relative light units (RLU) was not due to statistical error.

The activity of XBP-1s in this assay was confirmed using a KSHV RTA promoter luciferase construct, a gift from Dr. Sam Wilson. This reporter construct showed a positive response to both KSHV RTA and XBP-1s expression (figure 3.6.C). The expression of XBP-1s in the transformed cells used for luciferase assay was confirmed by using Western blot analysis (figure 3.6.A and B inserts).

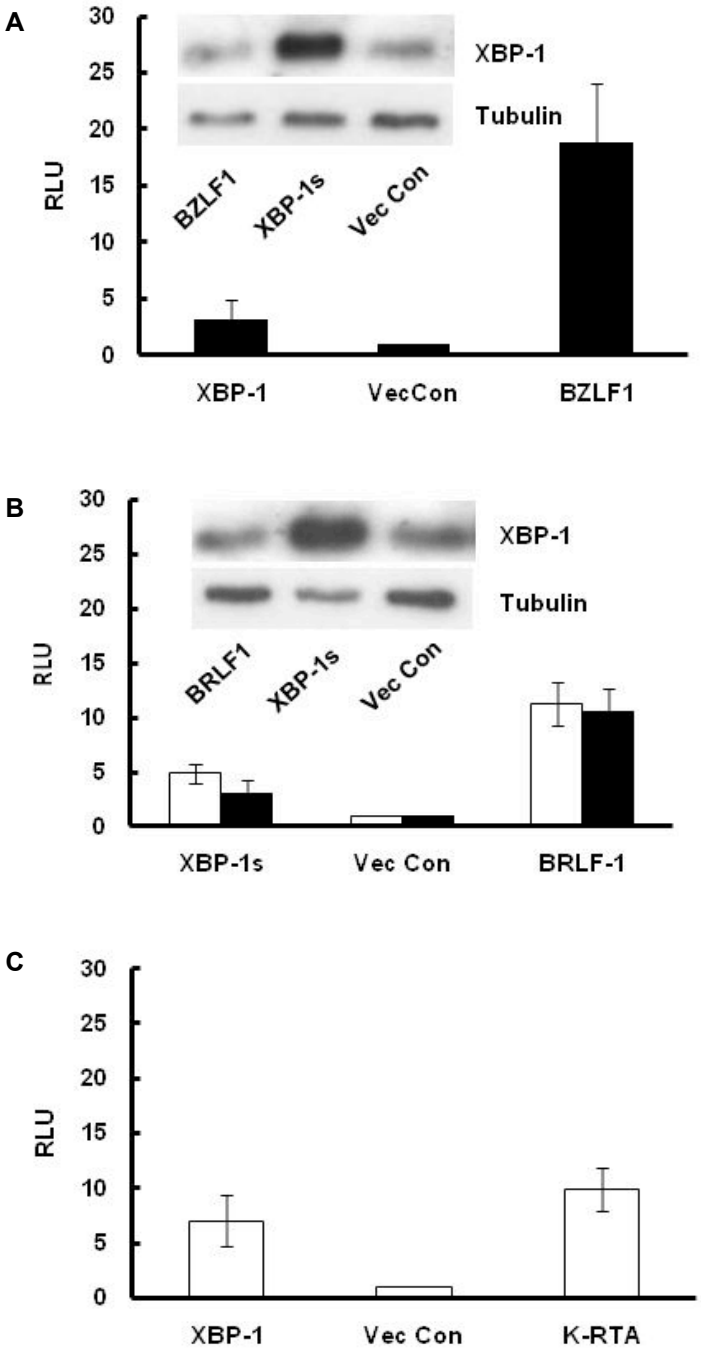


Figure 3.6. BZLF1 and BRLF1 promoter response to XBP-1s. RLU is the relative light units, indicating changes of luciferase activity compare with the vector control. The black bar represents methylated promoter luciferase construct, and the white bar represents unmethylated promoter luciferase construct. A) XBP-1s weakly transactivates the Zp in luciferase reporter assay in HEK 293 cells, although the weak response is statistically insignificant ($P>0.05$, two tailed T-test). BZLF1 over-expression transactivates the Zp as a positive control. B) BRLF1 and XBP-1s expression transactivate the Rp. The response is lower than the positive control, but statistically significant ($P<0.05$, two tailed T-test). The XBP-1s expression level in Zp (4A, insert) and Rp (4B, insert) transfected cells co-transfected with BZLF-1, BRLF-1 or XBP-1s was determined by Western blot. C) The K-RTA promoter was used as positive control for XBP-1s expression and activity. The promoter is transactivated significantly ($P<0.05$, two tailed T-test) by both K-RTA and XBP-1s. The results show the mean of three or more experiments with the error bars showing the standard deviation of these results.

3.2.7. XBP-1 is unspliced in different B-cell lymphomas, and splicing is not induced by TPA or VPA

XBP-1 is expressed retaining a 26 nucleotide (nt) intron (XBP-1u, unspliced), resulting in a frame shift preventing the translation of the active transcription factor. The intron is then spliced to allow the translation of the active form XBP-1s (XBP-1s, spliced) under ER stress or terminal differentiation into B-cells (Calfon et al., 2002). XBP-1u is expressed in both PEL and BL cell lines.

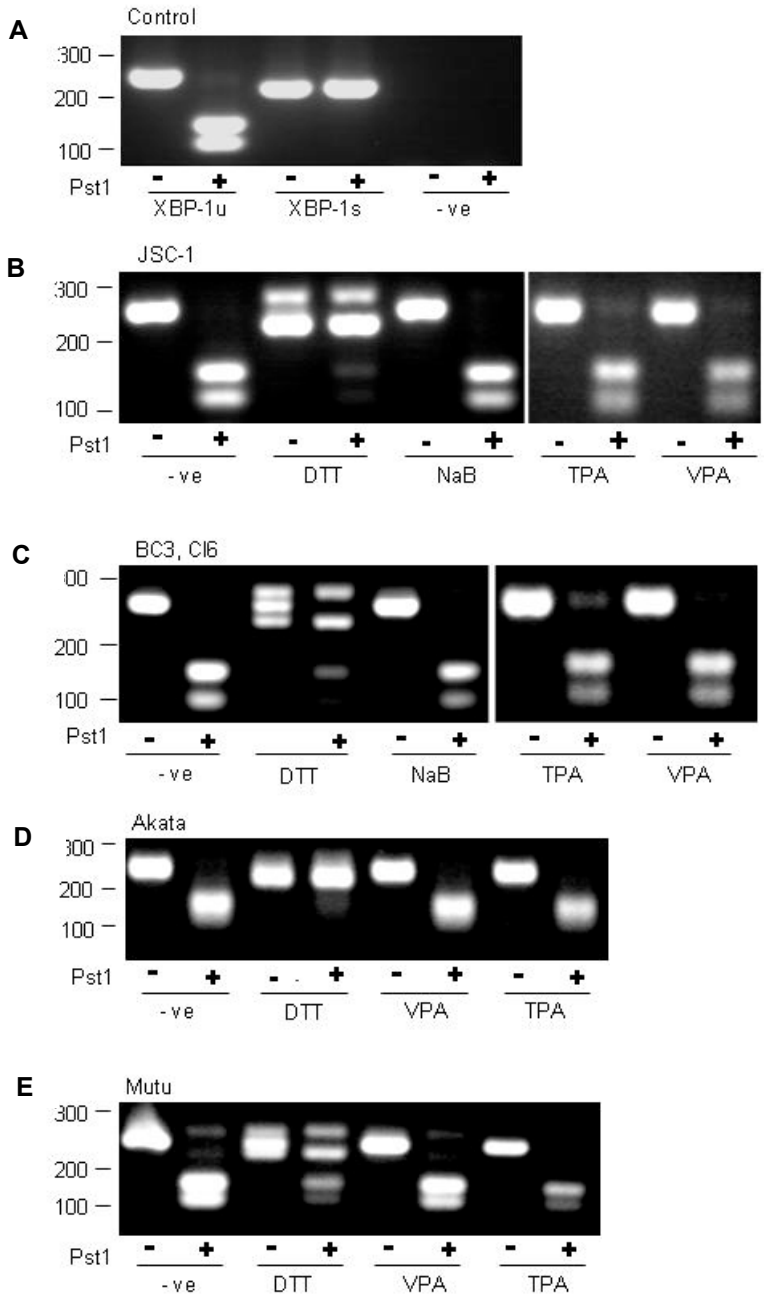
Previously, Wilson *et al* (2007) had shown that DTT, a reducing agent, is capable of inducing the KSHV lytic cycle in HEK 293 cells, and causing splicing of inactive XBP-1u mRNA into active XBP-1s mRNA by promoting the unfolded-protein response (UPR). As shown in the previous experiments, TPA can induce both KSHV and EBV lytic expression (shown by expression of KSHV RTA and BZLF1) (figure 3.3). Histone deacetylase inhibitors such as sodium butyrate (NaB) have also been shown to be able to induce EBV lytic cycle, although this is cell-line dependent (Countryman *et al.*, 2008). We set out to determine whether when the reducing agent DTT induced EBV lytic cycle it did so by causing XBP-1 splicing.

Unspliced XBP-1 is expressed in the two PEL cell lines: JSC-1 and BC3 cl6. Only DTT caused XBP-1 splicing, TPA, NaB and VPA did not cause XBP-1 splicing, even though TPA induces both KSHV RTA and BZLF1 expression (figure 3.7.B. C and F). Splicing of XBP-1 was also examined in BL cell lines such as Akata and Mutu I. XBP-1 is expressed in BL cell lines as an unspliced, inactive form. Again, only DTT caused XBP-1 splicing, both TPA and VPA cannot induce XBP-1 splicing (figure 3.7.D. E). Therefore, XBP-1 splicing does not induce EBV lytic reactivation in BLs. VPA and TPA induce reactivation of EBV lytic cycle, but not *via* XBP-1 splicing.

3.2.8. Induction of XBP-1 splicing by the UPR does not induce the KSHV and EBV lytic cycle.

To further examing the role of DTT in inducing XBP-1 slicing in both PEL and BL cell lines, we examined the ability of DTT to induce KSHV and EBV into their lytic cycles using Western blot for KSHV-RTA and EBV BZLF1. DTT resulted in significant cell death in PEL cells. (figure 3.7.H); after treating JSC-1 cells with

DTT for 10 minutes, the cells lose membrane integrity. Using trypan blue staining, 48 hours after treating the cells with DTT, more than 50% cell death was observed. Because of the toxicity of DTT on PEL cells, it is not possible to determine the effect of XBP-1 splicing and EBV/KSHV reactivation in PELs (figure 3.7.H). However, the BL cell line Akata is more tolerant of DTT treatment. DTT treatment of BL cells induced XBP-1 splicing (figure 3.7.D. E); however the ability of DTT to cause XBP-1 splicing in BL cells did not result in induction of BZLF1 protein expression (figure 3.7.G).



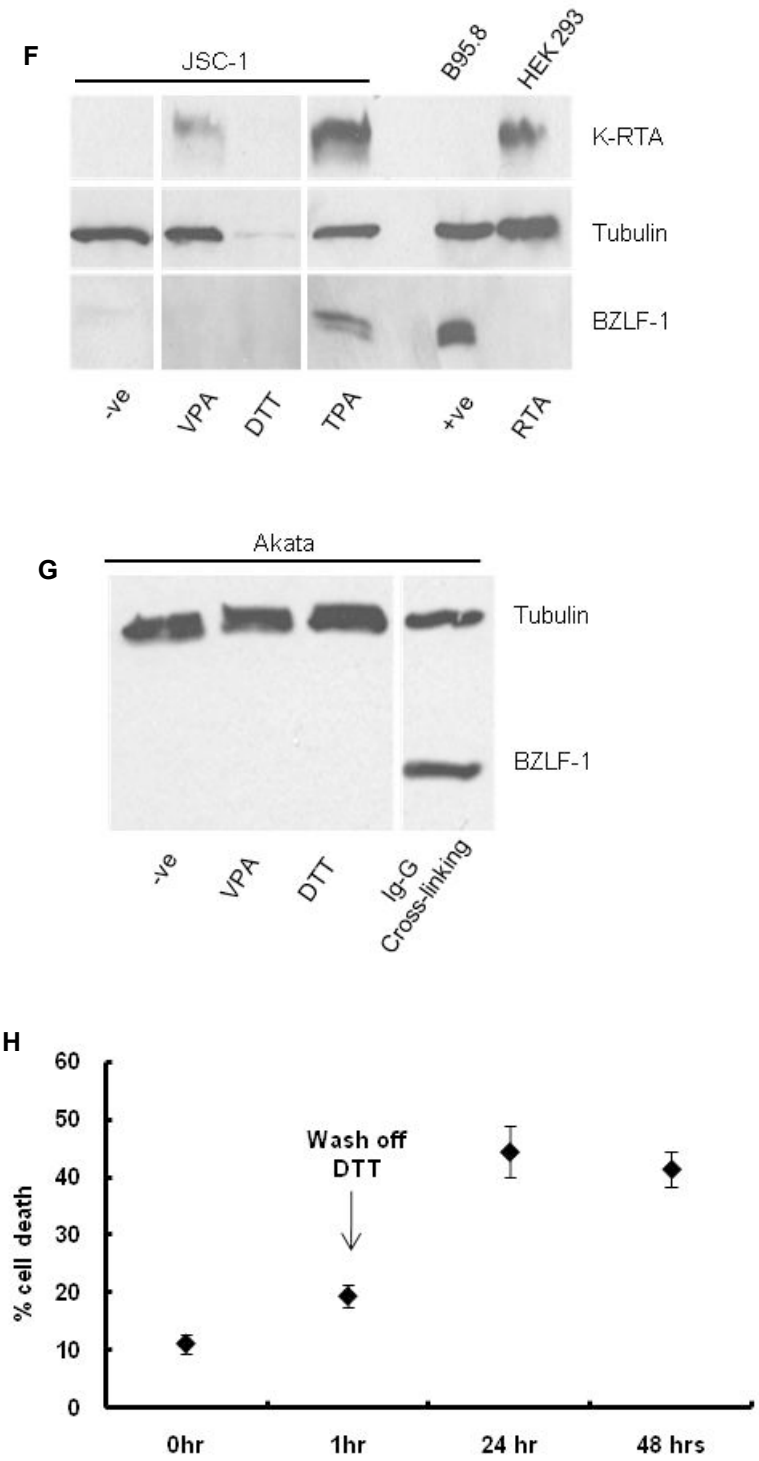


Figure 3.7. NaB, TPA and VPA do not induce XBP-1 splicing in PEL and BL cell lines. A) XBP-1 is expressed as the unspliced (XBP-1u) form, but the spliced active form (XBP-1s) can be detected following ER stress induction by resistance to digestion by the Pst1 restriction enzyme of the XBP-1s PCR product. Treating the cells with DTT induces ER stress and XBP-1u splicing in all lymphoma cell lines (B-E). However, NaB, TPA and VPA treatment do not induce XBP-1 splicing in the lymphoma cell lines (B-E) even though VPA induces K-RTA expression in the PEL JSC-1 (F) and TPA induces both K-RTA and BZLF1 in JSC-1 cells (F). VPA and DTT do not induce BZLF1 expression in the BL line Akata (G). DTT toxicity (H) results in PEL JSC1 cell death and the absence of detectable proteins (F).

3.2.9. Knocking down XBP-1s in BL cells

McDonald *et al.* (2010) showed that cross-linking B-cell receptors (BCR) resulted in XBP-1u splicing within 10 minutes. As BCR cross-linking of BL cell line induces the EBV lytic cycle (figure 3.4 lane 4). We decided to knock down XBP-1s in BL cell lines and investigate the effect it had on EBV reactivation following BCR cross-linking. An shRNA construct targeting XBP-1 and luciferase had previously been used to knock down the expression and an irrelevant control respectively (Dalton-Griffin *et al.*, 2009; Wilson *et al.*, 2007). After transducing the Akata cell line with both hairpin lentiviruses, a Western blot analysis was performed to determine the level of XBP-1 protein expression in wild type, XBP-1 knock down (XBP-1 KD) and luciferase knock down (Luc KD) Akata cell lines (figure 3.8.A). No significant differences in the amount of XBP-1 protein expression in the three cell lines were detected. However, these results could be due to the low transduction rate of Akata cells, resulting in incomplete knock down. To further investigate, we cross-linked the BCR to induce EBV lytic reactivation. As shown in figure 3.8.B, after 1 hour cross-linking the BCR, wild type Akata cells, in comparison to both XBP-1 KD and Luc KD cells, expressed less BZLF1. However, 4 hours after BCR cross-linking (figure 3.8.C), BZLF-1 expression changed between the wild type and knock down cells; with high BZLF1 expression in wild type but not in knock down cells. However no difference in BZLF1 expression between the XBP-1 KD and Luc KD cells was seen suggesting the effect was not due to knocking down of XBP-1, but due to the transduction with shRNA lentiviral vectors. Since there is no difference between XBP-1 protein expression in the XBP-1 KD, wild type and Luc KD control, it is not possible to determine if BCR cross-linking induces EBV lytic cycle *via* XBP-1.

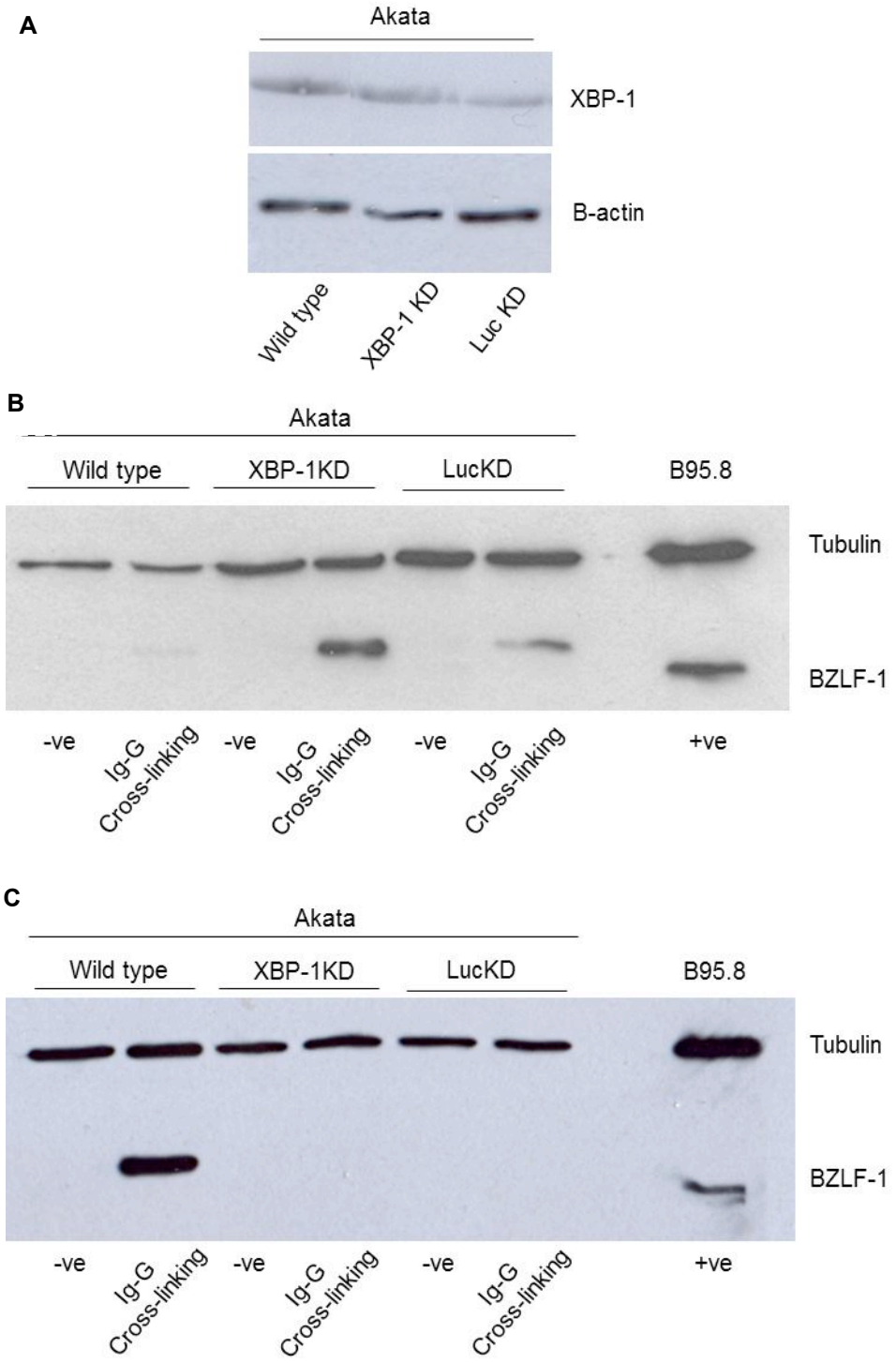


Figure 3.8. XBP-1 Knock down does not show different results compared to the irrelevant knocking down control. A) XBP-1 expression is slightly lower in XBP-1 KD akata cells, compare to wild type. However, it does not differ from the irrelevant control of luciferase knock down. B) one hour of BCR cross-linking with anti-IgG. BZLF1 expression is induced faster in XBP-1 KD Akata cells, compare to wild type. C) four hours after BCR cross-linking, BZLF-1 expression has been induced in wild type Akata cells, however, in both XBP-1 and Luc KD cell there is no BZLF1 expression.

3.2.10. Over-expression of XBP-1s or chemical agents do not induce expression of BZLF1 and BRLF1 mRNA in PEL or BL

Although XBP-1s over-expression did not induce detectable BZLF1 and BRLF1 protein expression in PEL or BL cell lines, the promoter based luciferase assays showed weak transactivation response between XBP-1s and Zp and Rp in HEK 293 cells. Therefore, we used quantitative real-time PCR (Q-RT-PCR) to determine the mRNA level of BZLF1 and BRLF1 in both PEL and BL cell lines in response to XBP-1s over-expression. The effects of various chemical inducers of the KSHV and EBV lytic cycles on the mRNA level of BZLF1 and BRLF1 were also determined in this experiment.

In Akata cells, the BZLF1 and BRLF1 mRNA level did not increase after XBP-1s over-expression, nor following TPA, VPA or DTT treatment (figure 3.9.A). BCR cross-linking with anti-human IgG was used as positive control and BZLF1 mRNA level increased significantly ($p<0.05$, two tailed T-test). However, despite the increase of BRLF1 mRNA level after BCR cross-linking, this was not statistically significant ($p> 0.05$, two tailed T test). This is consistent with the lack of BRLF1 protein expression (figure 3.5.A), and also a previous study (Amon et al., 2004) where BCR cross-linking only transactivated Zp, but not Rp.

The same experiments were carried out in JSC-1 cells. The EBV BZLF1, BRLF1 and KSHV K-RTA mRNA expression were determined using Q-RT-PCR (figure 3.9.B). After XBP-1s over-expression, or VPA or DTT treatment, the expression level of BZLF1 and BRLF1 mRNA did not increase significantly ($p>0.05$, two tailed T-test). The K-RTA mRNA level increased significantly after overexpressing XBP-1, DTT or VPA treatment. TPA is used as positive control, and compared to the negative control, the BZLF1 and K-RTA mRNA level both increased significantly ($p<0.05$, two tailed T-test) consistent with ZTA and K-RTA protein expression (figure 3.5.B). These data confirmed that XBP-1s alone does not transactivate the EBV Zp or Rp in PEL or BL cell lines.

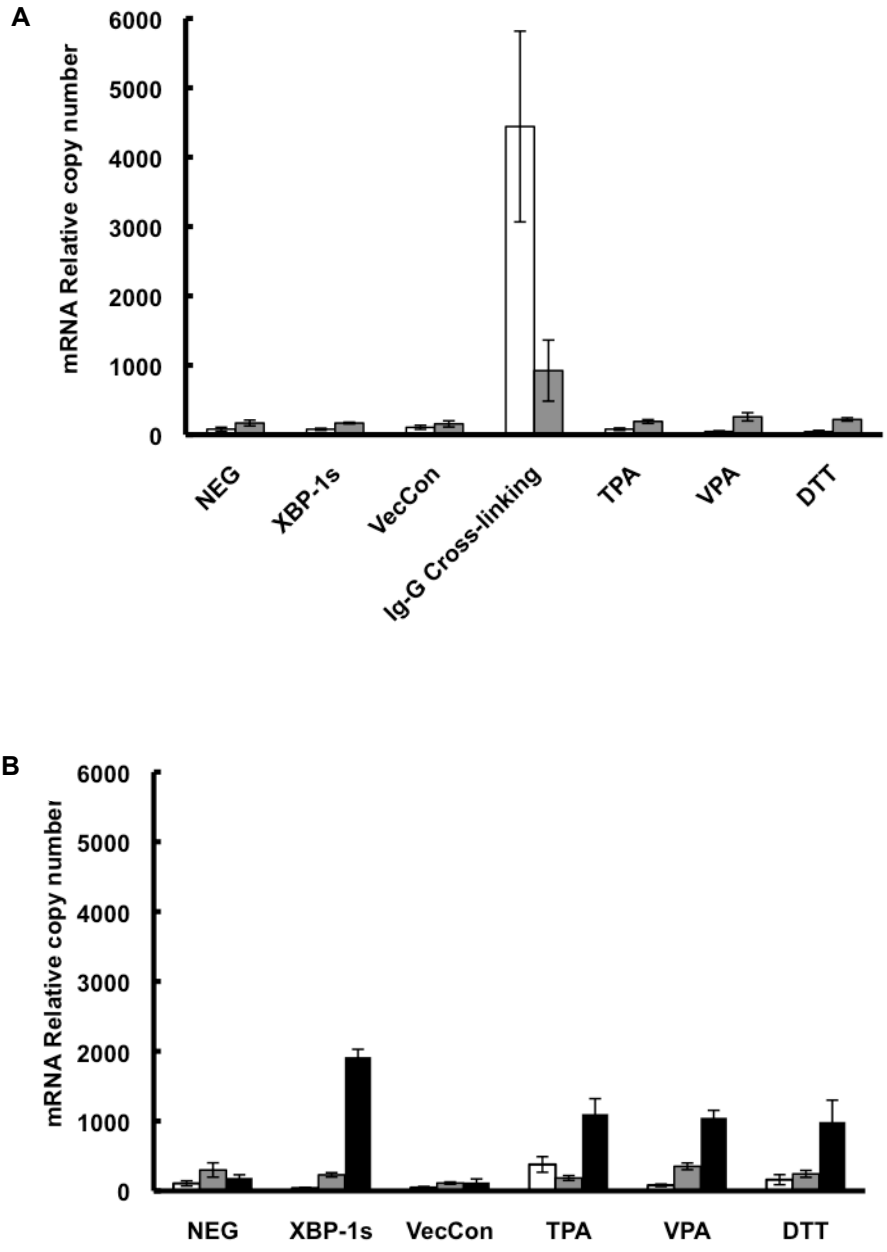


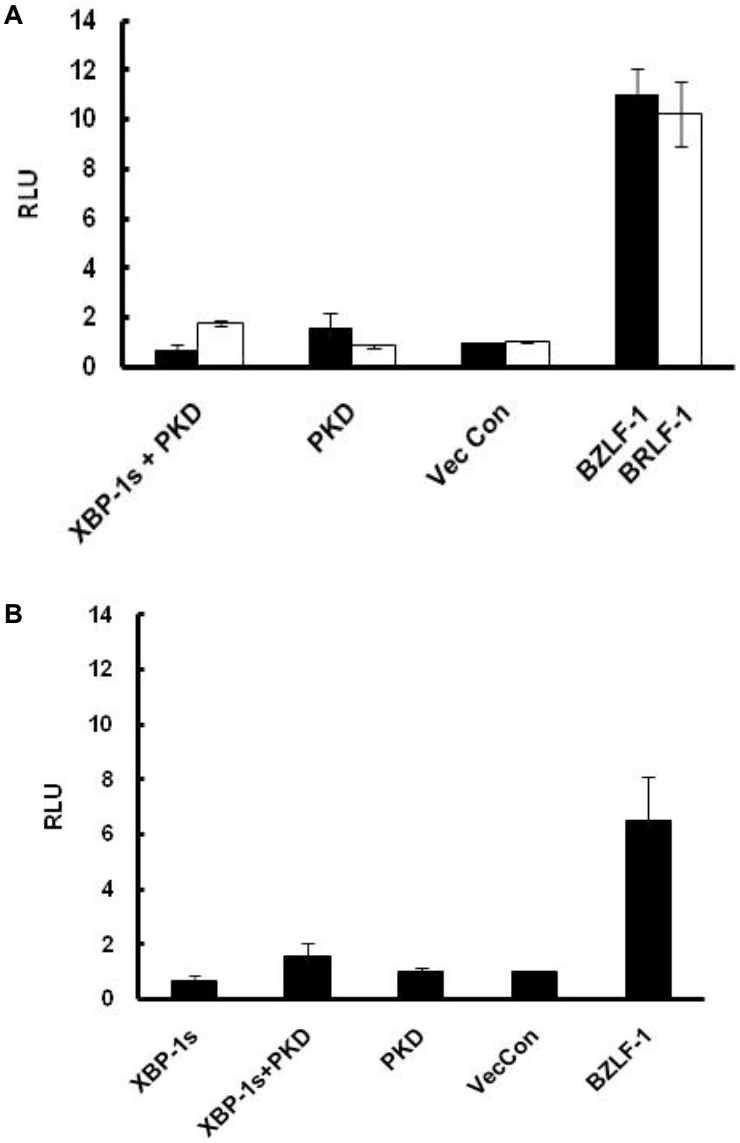
Figure 3.9. Over-expression of XBP-1s in PEL and BL cell lines does not increase the mRNA expression level of BZLF1 and BRLF1. Quantitative real time PCR was used to determine the expression of BZLF1 mRNA (open bars), BRLF1 mRNA (grey bars) and K-RTA mRNA (black bars) after transduction with XBP-1s lentivirus or treatment with chemical inducers for A) Akata and B) JSC-1 cell lines. A two-tailed T-test was performed to determine the possibility of false results. A) Neither XBP-1s over-expression or TPA, VPA or DTT treatment was able to increase BZLF1 or BRLF-1 mRNA expression. BCR cross-linking induced BZLF1 mRNA expression significantly compared to the vector control ($P>0.05$, two tailed T-test). B) In JSC-1 cells expression of XBP-1s only increased K-RTA mRNA expression. TPA, VPA and DTT also increased the K-RTA mRNA expression significantly ($P<0.05$, two tailed T-test). TPA treated JSC-1 cells expressed BZLF1 mRNA compared with negative control significantly ($P<0.05$, two tailed T-test). The two tailed T-test analysis indicates that the increase of mRNA levels, as described above, are due to either the effect of overexpressing XBP-1s or chemical treatment.

3.2.11. XBP-1s and Protein Kinase D together do not induce BZLF1 expression in PEL cell lines

Previous results showed that XBP-1s alone does not transactivate the EBV BZLF1 promoter and therefore does not induce the EBV lytic cycle in either PEL or BL cell lines. Other studies indicated that despite the inability of XBP-1s to induce EBV into the lytic cycle that, in combination with protein kinase D, XBP-1s was able to induce EBV BZLF1 expression (Bhende et al., 2007).

We used the luciferase assay to examine the effect of PKD in combination with XBP-1s on Zp, Rp and K-RTA promoter activity. A constitutively active PKD (pPKDm-IG) was used in the study. In HEK 293 cells, XBP-1s and PKD combined together weakly transactivated Rp but not Zp (figure 3.10.A). In order to ensure that this result was not cell line specific, we also carried out the BZLF1 promoter luciferase assay in HeLa cells. XBP-1s and PKD together weakly transactivated the Zp in HeLa cells (figure 3.10.B). However the weak increases in promoter activity in both HEK 293 cells and HeLa cells were not statistically significant ($P>0.05$, two tailed T-test). BZLF1 and BRLF1 were both able to transactivate their respective promoters in both cell types.

Over-expressing either PKD or XBP-1s alone or together in JSC-1 cells also did not induce BZLF1 and/or BRLF1 protein expression (figure 3.10.C), which confirmed the luciferase assay. K-RTA protein expression was induced whenever XBP-1s was over-expressed, with or without PKD. This indicates that PKD does not affect transactivation of K-RTA by XBP-1s. PKD alone did not induce K-RTA expression. Using Q-RT-PCR we further confirmed that the BZLF1 and BRLF1 mRNA level did not increase when PKD was expressed alone or in combination with XBP-1s (figure 3.10.D). K-RTA mRNA expression increased significantly whenever XBP-1s was overexpressed ($P<0.05$, two tailed T-test) (figure 3.10.D). Overexpressing PKD prior to XBP-1s also increased the K-RTA mRNA level more than XBP-1s expression alone or at the same time as PKD. Therefore, over-expression of PKD alone does not induce EBV BZLF1, BRLF1 and KSHV K-RTA expression. Additionally, over-expression of active PKD and XBP-1s does not induce EBV BZLF1 and BRLF1.


A
RLU
14
12
10
8
6
4
2
0
XBP-1s + PKD
PKD
Vec Con
BZLF-1
BRLF-1
B
RLU
14
12
10
8
6
4
2
0
XBP-1s
XBP-1s+PKD
PKD
VecCon
BZLF-1

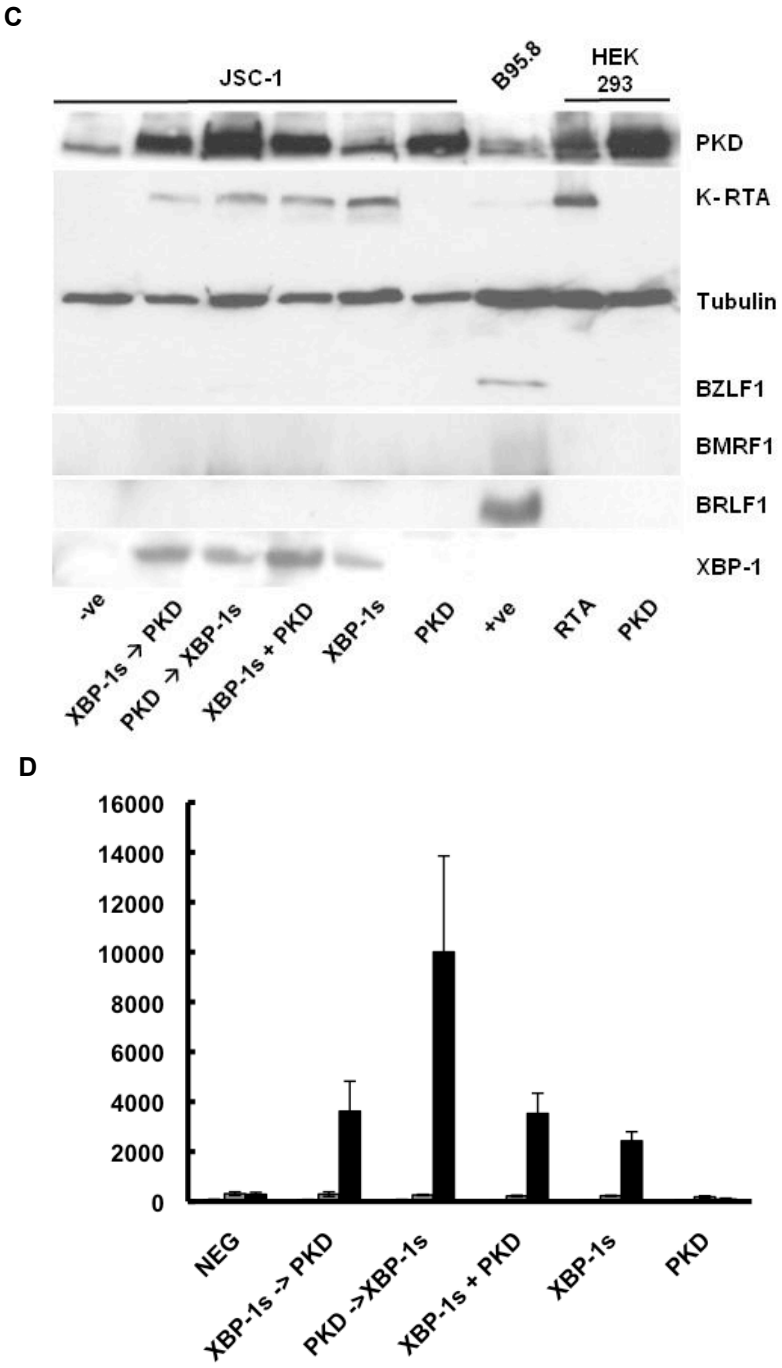


Figure 3.10. Co-expression of XBP-1s and active PKD does not transactivate Zp or induce BZLF1 expression. RLU indicates the relative light unit of luciferase activity. A) HEK 293 cells were transfected with Zp (black bars) or Rp (open bars) luciferase reporter vectors with either active PKD or XBP-1s together with active PKD. Active PKD alone shows weak Zp activity and together with XBP-1s weak Rp activity compared to the control empty vector (VecCon) and BZLF1 or BRLF1 expression vectors. However, the increase of luciferase activity was not statistically significant compared to the vector control ($P>0.05$, two tailed T-test) B) The Zp promoter assays were repeated in HeLa cells where only XBP-1s and PKD together showed weak promoter activity compared to BZLF1 over-expression. However, this was not statistically significant compared to the vector control ($P>0.05$, two tailed T-test). C) Transduction of JSC-1 cells with XBP-1s and active PKD either alone, in combination, or sequentially (arrows) does not induce BZLF1, BRLF1 and BMRF1 expression in PEL, but all combinations that include XBP-1s induce K-RTA expression. D) The mRNA expression, measured by quantitative real time PCR after XBP-1s and/or active PKD transduction for BZLF-1 mRNA (open bars), BRLF1 mRNA (grey bars) and K-RTA (black bars). Only K-RTA mRNA expression increased significantly ($P<0.05$, two tailed T-test) after XBP-1s transduction with or without PKD expression. Two tailed T test analysis was performed to determine the possibility of the results being statistically significant.

3.2.12. Examing genetic variation in the Z promoter sequences from mature B-cell lymphomas and wild type EBV

The sequence of the Z promoter in both PEL and BL cell lines was not known. We therefore sequenced the promoter region of these cell lines to determine if polymorphism in the promoter could affect the ability to induce BZLF1 expression. We also sequenced the Zp region from the luciferase reporter construct to ensure that the sequence did not differ from that found in the cell lines. The Zp sequence was compared with the EBV type 1 and type 2 (Genbank ref. NC_007605 and NC_009334 respectively) Zp region from the Genbank database.

There are three nucleotide differences between the EBV type 1 and type 2 in the Zp sequences (figure 3.9). The Z promoter luciferase construct that was used originated from the B95.8 cell line. The Zp region, from B95.8, Mutu I and Raji cell lines, were all identical to the Zp sequence of type I EBV; Akata, HBL6 and have this type 2 EBV Zp sequence.

A previous study carried out by Sun and Thorley Lawson (2007) indicated that, though computational prediction, Zp contains two putative XBP-1 binding motifs designated Zp1 and Zp2, at the ZII and ZID regions of the Zp (figure 3.11), with GTCA and ACGT motifs respectively. They have also shown that XBP-1s only binds to the Zp1 but not to the Zp2 region. However, using computational prediction software Jasper, ACGT is predicted to be the binding site for XBP-1s. In addition, previous studies using MD scanning program have also predicted ACGT to be an XBP-1s binding site (Acosta-Alvear *et al.*, 2007). As mentioned earlier in this chapter, XBP-1s is able to bind and transactivate KSHV immediatel early protein K-RTA. It has been shown that XBP-1s binds to the XRE element in the K-RTA promoter, containing the ACGT motif (Dalton-Griffin *et al.*, 2009). It is unclear why XBP-1s does not bind to the ACGT motif in the Zp.

There are three single nucleotide differences between the type 1 and type 2 EBV Zp sequences. However, the three single nucleotide base differences are not located in either the Zp1 or Zp2 regions. Therefore the lack of response to XBP-1s seen in the Zp is not caused by mutation in the Z promoter.

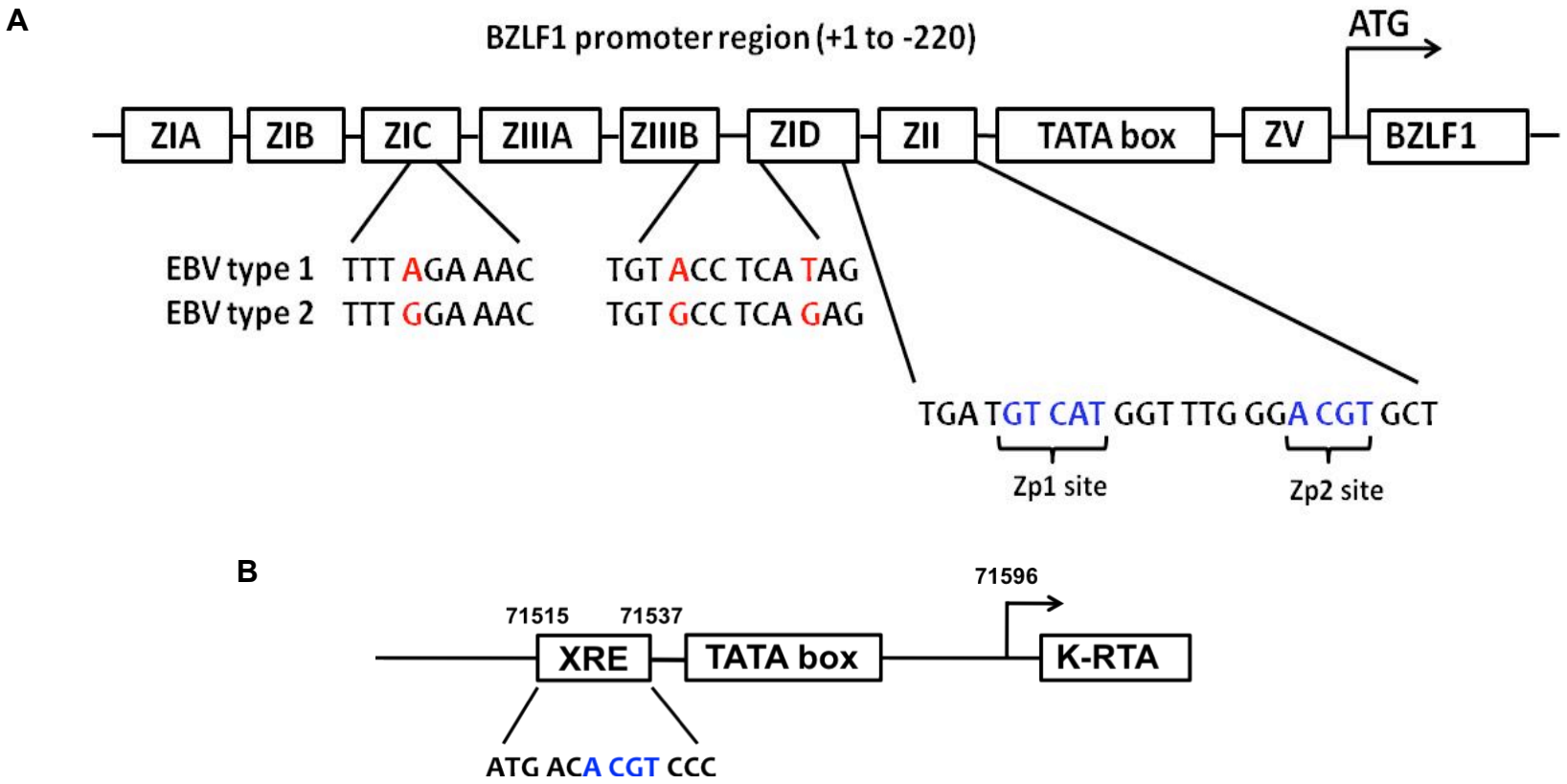


Figure 3.11. Schematic diagram of the a) BZLF1 promoter Zp, with the regulatory elements; and B) K-RTA promoter. A) The nucleotide sequences of the ZIC and ZIIIB/ZID region, indicating the difference between EBV type 1 and type 2. The nucleotide sequence of the ZID/ZII region is also shown, which contains the two putative XBP-1 binding motifs Zp1 and Zp2. Zp2 contains ACGT motif. This diagram indicated that the difference in nucleotide sequence between EBV type 1 and type 2 should have no effect in interacting with XBP-1s. B) XBP-1 response element (XRE) in the K-RTA promoter contains ACGT binding motif. Computational program JASPER has been used to determine the binding motif for XBP-1, which contains the sequence ACGT.

3.3. Discussion

Plasma cell terminal differentiation and the associated expression of XBP-1s are linked to EBV and KSHV lytic cycle induction (Bhende *et al.*, 2007; Laichalk & Thorley-Lawson, 2005; Sun & Thorley-Lawson, 2007; Wilson *et al.*, 2007; Yu *et al.*, 2007a). Recently this has been extended beyond human gamma herpesviruses with the demonstration that plasma cells account for most of the lytic reactivation of murine herpesvirus 68 in mice (Liang et al., 2009). It is known that XBP-1s transactivates and induces the expression of KSHV IE protein K-RTA in PEL cells, which initiates the rest of KSHV lytic cycle (Wilson *et al.*, 2007). However, the role of XBP-1s in EBV lytic cycle induction is less well defined. In this chapter, we investigate the mechanism that underlines the initial induction of KSHV RTA and EBV BZLF1. XBP-1s alone induces KSHV into lytic cycle by transactivating the KSHV RTA promoter (ORF50 promoter) in PEL cell lines. However, this pathway of reactivation does not induce EBV BZLF1 and BRLF1 protein expression in either PEL or BL cell lines. Over-expressing XBP-1s also does not induce the production of BZLF1 or BRLF1 mRNA. Additionally, XBP-1s does not transactivate BZLF1 promoter in either HEK 293 cells or HeLa cells. Furthermore, XBP-1s and constitutively activated PKD together do not induce BZLF1 protein expression or mRNA production. The BZLF1 promoter sequences between different cell lines contain no mutations that could affect the transactivation. These results suggest that XBP-1s does not activate EBV lytic cycle in either PEL or BL cells, despite being able to activate KSHV lytic cycle in PELs

Previous reports have shown various effects of XBP-1s in EBV reactivation. Sun and Thorley-Lawson (2007) have shown that XBP-1s is able to transactivate BZLF1. Bhende *et a*l. (2007) have shown that XBP-1s alone does not transactivate BZLF1 but can in combination with PKD expression, although the mechanism of reactivation is unclear. However, these reports did not carry out experiments in either a mature B-cell background, or used murine XBP-1s that differ from human XBP-1s. Sun and Thorley-Lawson (2007) have carried out studies in an multiple myeloma cell line MM1.S, which constitutively express XBP-1s. Sun *et al*. (2007) have shown that in HeLa cells XBP-1s cannot transactivate Zp; whereas Bhende *et al*. (2007) showed, also in HeLa cells, that XBP-1s weakly transactivates Zp and Rp and the effect of Zp could be

enhanced by co-expression of constitutively active PKD. Our data in HEK 293 cells supports the observations that XBP-1s alone can weakly transactivate Zp and Rp. Similarly we showed co-expression of XBP-1s and PKD weakly transactivates Rp in HEK 293 cells although not to the magnitude described previously. This may reflect differences in assay sensitivity or may reflect the fact that we have used human XBP-1s in our studies rather than murine XBP-1s used by Bhende *et al.* (2007). Despite the weak transactivation of the Zp and Rp in these two cell lines, the relevance of these observations in HEK 293 and HeLa cells and the response of EBV to XBP-1s and PKD in B-cell tumour lines is questionable. We have also carried out the study on mRNA level and immediate early protein expression in B-cell lineages. More evidence has suggested that EBV undergoes different pathways of infection and reactivation in epithelial cells and B-cells. It is possible that the differences in results are also due to the type of cell lines used in the studies.

The ability of XBP-1s to transactivate either Zp or Rp most likely depends on both the cell type and on the nature of individual cell lines. This is supported by the observation that distinct chemical inducing agents have different effects on the induction of EBV and KSHV lytic cycles in various lymphoma lines (Countryman *et al.*, 2008; Miller *et al.*, 1997), suggesting that the reactivation of EBV by XBP-1s can vary depending on cell line. In a lymphoblastoid and an NPC cell line XBP-1s and PKD clearly activated EBV BZLF1 protein expression, but BL or PEL cell lines were not tested (Bhende et al., 2007). In an unusual multiple myeloma cell line latently infected with EBV, only low levels of BZLF1 transcript could be induced by XBP-1s. This observation contrasts with recent data indicating EBV can latently infect multiple myeloma cell lines *in vitro*, and in these circumstances the endogenous, active XBP-1s that is expressed by multiple myeloma cell lines, does not drive the EBV lytic cycle (Anastasiadou et al., 2009). Here we show that despite weak transactivation of Zp and Rp by XBP-1s in HEK 293 cells, there are no detectable levels of BZLF1 and BRLF1 mRNA or proteins after XBP-1s over-expression in B-cell tumour lines.

Why EBV Zp is refractory to XBP-1s in PEL and BL even when PKD is co-expressed is not clear, although the role of the chromatin structure on Zp promoter is likely to be important. Indeed, the binding of myocyte enhancer factor 2D (MEF2D) to Zp recruits class II histone deacetylases (HDACs)

presumably promoting a repressed chromatin structure on Zp in lymphoma cell lines (Gruffat *et al.*, 2002; McDonald *et al.*, 2010). It has been suggested that methylation and histone acetylation of Zp are involved in maintaining viral latency in BL (Bergbauer *et al.*, 2010; Flower *et al.*, 2011; Murata *et al.*, 2012). De-repression of Zp associated chromatin by PKD was suggested as the mechanism that allows XBP-1s to activate Zp (Bhende et al., 2007). PKD is a member of the serine/threonine protein kinase family (Rey et al., 2006) which can be activated by protein kinase C (PKC) (Zugaza et al., 1996), a cellular target of phorbol esters such as TPA (Wang, 2006). PKD can also be activated by B-cell receptor cross-linking *via* PKC pathway (Matthews *et al.*, 2000). In our hands TPA induces both K-RTA and BZLF1 in PEL and BCR cross-linking induces BZLF1 in BL. It indicates that both TPA and BCR cross-linking can activate PKD *via* PKC pathway. Activated PKD can phosphorylate and therefore inactivate class IIa HDACs, similar to the effect of HDAC inhibitors (HDACi), such as trichostatin A (TSA) (Chang and Liu 2000), valporic acid (Feng & Kenney, 2006) and sodium butyrate (NaB). As HDACi are also capable of inducing EBV lytic reactivation in various cell lines (Gradoville et al., 2002), we investigated if XBP-1s is also induced by any of these chemical inducers. Here we show that NaB, VPA or TPA does not induce activation of XBP-1s in PEL and therefore the effects of HDACi and TPA on the Zp is independent of XBP-1s.

BCR cross linking induces the EBV lytic cycle and can also transiently induce XBP-1s production. However, this transient XBP-1s production cannot unambiguously be linked to BZLF1 expression (McDonald et al., 2010). Here we used DTT to chemically induce XBP-1s in PEL and BL cell lines. Although DTT is able to induce XBP-1 splicing and induce K-RTA in an XBP-1s dependent manner in HEK 293 cells (Wilson et al., 2007), it causes significant cell death in PEL cells making it impossible to determine K-RTA, BZLF-1 or BRFL-1 protein expression. Nevertheless, K-RTA mRNA expression was detected following DTT treatment of PEL (Figure 6B) and DTT, which does not cause cell death in BL cells, induces XBP-1s production, but does not induce BZLF1 expression. Taken together these data suggest that induction of EBV lytic cycle in the lymphoma cell lines of PEL and BL is not linked to XBP-1s activity, although the effects in B-cells *in vivo* may be different.

In the dually infected PEL cell backgrounds, other studies showed that a selective induction of one viral lytic cycle cross represses the other viral lytic cycle and co-expression of both KSHV and EBV immediate early proteins results in mutual inhibition (Jiang et al., 2008; Miller et al., 1997; Xu et al., 2007). This suggests a selective lytic switch that induces only one virus is required for successful EBV or KSHV lytic replication in PEL. Here we show that XBP-1s is able to induce KSHV lytic replication alone in either singly or dually infected PEL, but does not induce the expression of EBV BZLF1 and BRFL1 IE proteins or transcripts in PEL or BL. XBP-1s with PKD also does not induce the expression of BZLF1 and BRFL1 IE or transcripts in PEL. This indicates that XBP-1s specifically induces only KSHV into lytic reactivation even in the presence of EBV, therefore it is possible that XBP-1s is acting as the “switch” to induce reactivation of specific viruses.

Collectively, our results show that XBP-1s does not reactivate EBV lytic cycle in mature B-cells, with or without constructively activated PKD. The known reagents that can successfully induce EBV reactivation, such as TPA and HDACi, do not induce XBP-1 splicing. With recent study showing ER stress induces EBV lytic reactivation (Taylor *et al.*, 2011), and ER stress being part of the plasma cell differentiation, it is possible other factors involved in the ER stress pathway are playing important roles in EBV reactivation in B-cells.

Chapter 4

Results: Establishing methods for EBV whole genome sequencing

4.1. Introduction

Little is known about the genome variation of EBV, despite the high prevalence of EBV infection worldwide. There are currently five EBV genome sequences in Genbank of which two are from Burkitt's lymphoma cell lines; one from a marmoset cell line (B95.8), which was infected by EBV originating from an Infectious mononucleosis sample; and two from nasopharyngeal carcinoma (NPC) cell lines GD1 and GD2. Within these five genome sequences, there is only one EBV type 2 genome sequence. The reference genome of EBV type 1 is an artificial fusion between the B95.8 genome sequence with the deletion filled in by the Daudi (BL) cell line EBV sequence. There is no primary isolate wild type (WT) EBV whole genome sequence. WT EBV whole genomes in this study indicate the EBV whole genomes which are not associated with transformed B-cells. Additionally, there is no whole genome sequence of EBV present in dually infected PEL cell lines, or from other diseases that is associated with EBV.

It is well established that the genomes of RNA viruses vary extensively. It is also established that genome variation is less extensive for DNA viruses. However, a recent study has shown that Human cytomegalovirus (HCMV), a member of the β-herpesvirus family, is more variable at both the genome and amino acid level than expected (Renzette *et al*., 2011). Previous studies carried out by Kelly *et al*., (2002) have showed that in BL samples, both cell lines and primary clinical samples, contain EBV genome variants known as Wp-restricted EBV that contain a deletion in the EBNA2 gene of the genome. More studies for another herpesvirus, varicella zoster virus (VZV), also showed variation between viruses. Samples taken from volunteers who had been vaccinated against the same strain of VZV vaccine have shown varied VZV genomes compared with each other and the vaccine strain (Breuer *et al*., 2007).

Due to the small number of EBV genome sequences that are available currently, it is difficult to determine the extent of genetic variation in the EBV

genome either between infected people, within a person over time, or in different diseases associated with EBV. Here we establish the sequence of EBV from different clinical sources using Next Generation Sequencing and methods to assemble whole EBV genomes.

4.2. Results

4.2.1. Assays for measuring EBV viral load (qPCR)

To date sequencing large DNA virus genomes has involved either sequencing λ phage genome libraries, bacterial artificial chromosomes or multiple PCR products covering the whole genome. With large viral genomes such as herpesviruses, the quantity of PCR products required to cover the whole genome became too large, increasing costs and time and also increasing the possibility of errors during the process.

Sequencing herpesvirus genomes by Next Generation Sequencing is also difficult due to the small amounts of EBV DNA relating to the host DNA. Direct deep sequencing from patient samples, such as the tumour tissue from a NPC patient (Liu *et al.*, 2011), resulted in 94% of the reads matching human DNA and 0.027% of the reads matching viruses and, within that, only 17% (or 0.0142% of total reads) matching to EBV. With these small percentages of reads that are positive to EBV, it is both time consuming and expensive to sequence EBV genomes from clinical samples, which prevents sequencing EBV whole genomes in large numbers. To routinely sequence EBV using high throughput Next Generation Sequencing therefore requires samples with high EBV genome copy number.

In order to determine the amount of EBV in the samples, a quantitative PCR (qPCR) assay was established and used prior to the sequencing to determine the EBV genome copy number in the samples. Primers and probes were designed against the EBNA-LP gene. The sensitivity of this assay allowed the detection of EBV genome copy to 10^2 genomes. GAPDH is used to determine amount human DNA in the samples. The EBV qPCR assay was linear over four orders of magnitude of EBV copy number (figure 4.1). The qPCR of GAPDH is more efficient in comparison to the EBV qPCR, since the gradient of GAPDH standard curve is about 30% higher than the EBV qPCR standard curve, indicating the ability to detect the gene more efficiently.

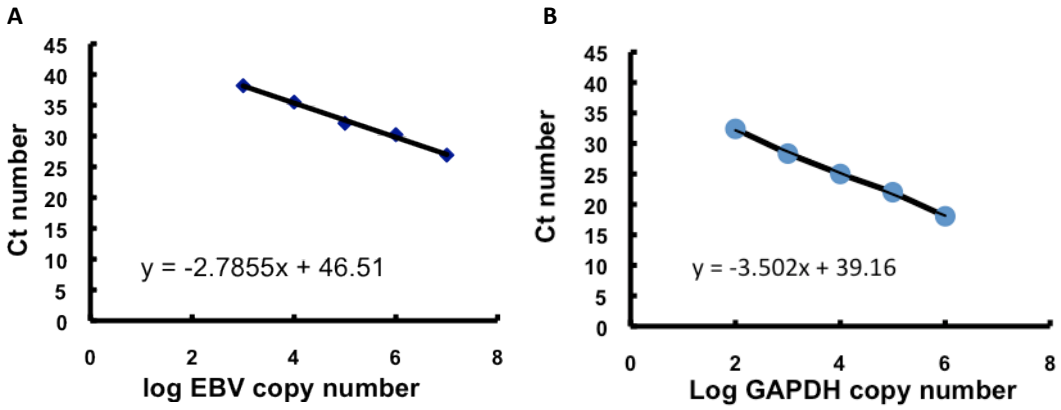


Figure 4.1. Standard curve of A) EBV B) GAPDH quantitative PCR. The primers and probes of EBV qPCR target the EBNA-LP region. Each point has been carried out in duplicate while setting up the assay. This qPCR standard curve had been carried out in duplicate when quantification of EBV and GAPDH genome copy is required during the study. The copy number of EBV and GAPDH genome in different samples has been calculated using the equation inserted in the figure. Y is the Ct number from the qPCR result and X is the log EBV or GAPDH copy number.

4.2.2. Identifying the source of wild type (WT) EBV sample used for sequencing

To date, the EBV genome sequences that are in Genbank are all from tumours associated with transform cells, with majority from either NPC or BL. However, the vast majority of EBV infections are asymptomatic and in addition to NPC and BL, EBV infection has been associated with diseases such as glandular fever or post transplant lymphoproliferative disease (PTLD). Sequencing an EBV whole genome not associate with any type of tumour would increase the understanding of the virus genome. Polymorphisms have been associated with aggressiveness of tumour progression. An EBV genome with no tumour association can provide information to determine whether any polymorphisms in EBV are associated with tumour formation.

EBV remains latent in memory B-cells, and transmits *via* saliva. Previous studies showed positive EBV detection in samples such as blood, buccal swap and saliva (Decker *et al.*, 1996; Hadinoto *et al.*, 2009). It is unclear which type of samples contain maximum EBV copy number for use in deep sequencing. Table 3.1 showed the EBV copy number from different sample types. Samples from two volunteers showed an undetectable (UD) amount of EBV. With one

volunteer showing the presence of EBV in all three types of samples, it is clear that saliva contains the highest level of EBV. Moreover, it has been shown that EBV is shed *via* saliva constantly through time with high copy number (Hadinoto *et al.*, 2009). Although no EBV presence as detected in the saliva samples from volunteers 1-3, it is possible that the qPCR sensitivity is not sufficient to detect EBV from these three samples. The primers used for the qPCR were unable to detect EBV when there was fewer than 10^2 copies per μl. Also, the method of collecting and processing saliva, such as if the cells have been removed prior to processing, could also affect the EBV copy numbers in these samples.

Table 4.1. EBV copy number in different types of sample

	PBMC	**Buccal Swap**	**Saliva**
Volunteer 1	UD	UD	UD
Volunteer 2	UD	UD	UD
Volunteer 3	3952	UD	UD
Volunteer 4	835	3947	7.6×10^6

Further saliva samples were obtained from 26 volunteers and EBV genome load was measured using qPCR (table 4.2). 16 samples contained detectable amounts of EBV. The samples with more than 10^6 EBV copies per μl (highlighted in table 4.2) were directly sequenced as stranded Illumina libraries at the Wellcome Trust Sanger Institute (WTSI). However, the sequencing read pairs did not produce matches to the EBV genome. Rather all reads matched salmon sperm DNA, which as added as a carrier at DNA extraction. Further saliva DNA was collected and extracted with no carrier DNA and sequenced at WTSI. Despite there being greater than 10^6 copies of EBV genome in the sample, the Illumina sequencing still failed to detect any EBV reads. This is most likely due to high human DNA content of either host or other organisms such as bacteria, resulting in majority of the read pairs not matching EBV. To overcome this problem, EBV genome will therefore need to be either separated from other DNA, especially the host DNA; or enriched, before deep sequencing can take place.

Table 4.2. EBV genome copy number from 26 saliva samples

Sample number	Copy no. /µl	Sample number	Copy no. /µl	Sample number	Copy no. /µl
1	4996	10	9257	19	$3.9x10^6$
2	4310	11	UD	20	$1.1x10^6$
3	UD	12	**$4.1x10^8$**	21	794
4	UD	13	28825	22	1488
5	UD	14	UD	23	$5x10^4$
6	UD	15	UD	24	$9.9x10^4$
7	**$5.3x10^7$**	16	$3.7x10^6$	25	2317
8	UD	17	UD	26	132
9	UD	**18**	**$1.2x10^8$**		

4.2.3. Using Gardella gel to extract episomal genome from latently infected cell lines

To date, no EBV whole genome sequences have been produced from PEL which is dually infected with KSHV. EBV remains in latency in both PEL and BL cell lines, indicating the EBV episomal DNA associates with the human DNA in the nucleus. Gardella gels were used to separate EBV genome from human DNA. The Gardella gel was designed in 1984 (Gardella *et al.*, 1984) for detecting herpesvirus episomes in latently infected cells. Figure 4.2 shows an illustrated diagram of a Gardella gel. With two sequential layers of agarose gels, the cell samples are lysed in the first tier with the virus episomal and human linear DNA being size separated in the second tier.

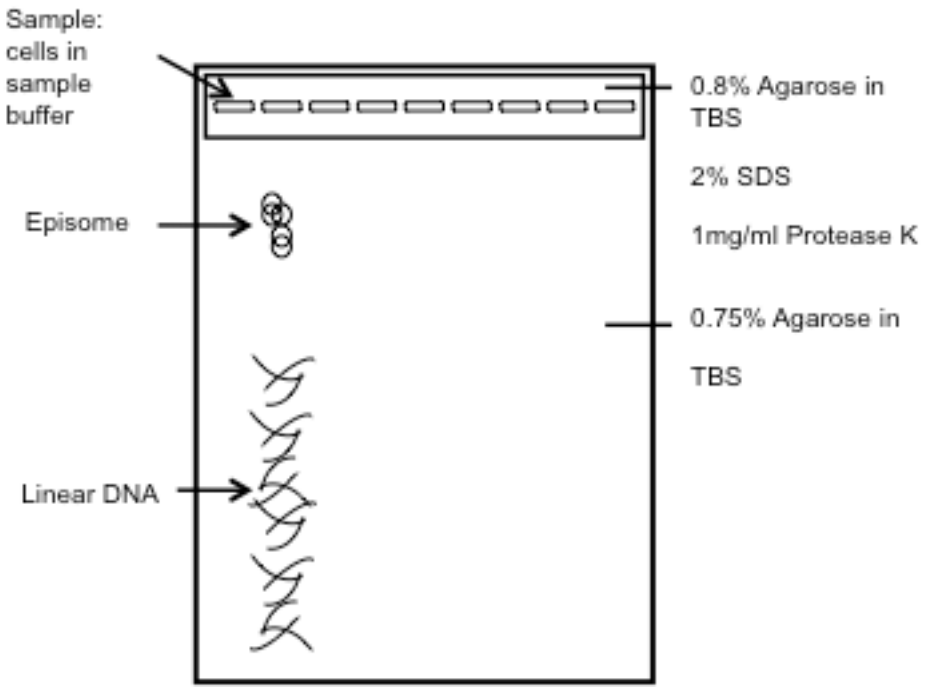


Figure 4.2. Schematic diagram of a Gardella gel. The Gardella gel separates the episomal and linear DNA from lysed cells. $1x10^6$ cells have been suspended in sample buffer and loaded into each well. Sections of the gel containing the episomal DNA were cut into 0.5cm pieces and soaked in distilled water to elute the DNA.

Here we adopted a method based on the Gardella gel used by Decter *et al* (1996), to separate the herpesvirus episomes from both JSC-1 and HBL-6 cell lines. Both of the PEL cell lines used are dually infected with KSHV and EBV, however this method cannot separate these two viral genomes. After gel electrophoresis, the gel was sliced into 0.5cm length sections, and the episomal DNA was eluted from the gel. The proximal distance from the well of episomal DNA was estimated according to previous literature (Decker *et al.*, 1996; Gardella *et al.*, 1984). PCR was performed to confirm EBV DNA was eluted from the gel (figure 4.3). The PCR result indicated that EBV genome could be detected spreading 6 cm from the loading well. Gel slices taken from other location of the Gardella gel were used as negative control. However the negative control also shows low level of EBV indicating the difficulty of estimating the discrete location of EBV within the gel. Also, it is unclear whether either the episomal or linear form of EBV genome is in the samples.

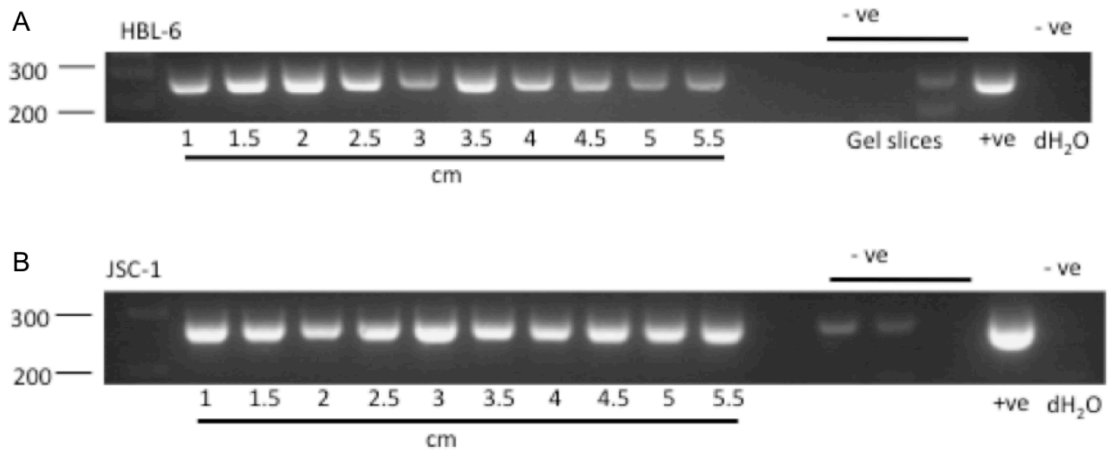


Figure 4.3. Determining the presence of the EBV genome after elution of the Gardella gel from A) HBL-6 and B) JSC-1 cell lines. Each PCR indicates the distance of the gel slice from the well. Random areas of the Gardella gel were also sliced and eluted and tested by PCR to ensure no contamination occurred during electrophoresis and the elution process, dH_2O was used as negative control for the PCR.

4.2.4. Using DNA polymerase phi29 with EBV specific primers to increase the viral copy number.

The total number of EBV genomes extracted from the cell lines using Gardella gels is between 10,000 and 23,000 copies, which is low in comparison to the other samples that had failed when sequencing directly from saliva. With 10^6 cells being loaded per well, and multiple episomes per cell, the recovery efficiency of EBV genomes from the Gardella gel is low. Phi29 DNA polymerase was used to increase the copy number of EBV with target specific primers. Phi29 is a proof-reading DNA polymerase, able to perform a rolling circle DNA amplification *in vitro* (Dean *et al.*, 2001). Since the DNA extracted from Gardella gel is episomal, Phi29 is ideal for increasing copy number with minimal introduction of DNA induced errors.

Initially, 6 EBV specific 20mers spaced evenly across the EBV genome were used in the PCR, which only increased the EBV copy number by 2-3 fold (figure 4.4). An additional 12 primers were added and the PCR repeated. This resulted in the EBV copy number increasing by 10 fold relative to input DNA. These samples were sent to WTSI for sequencing, but again failed.

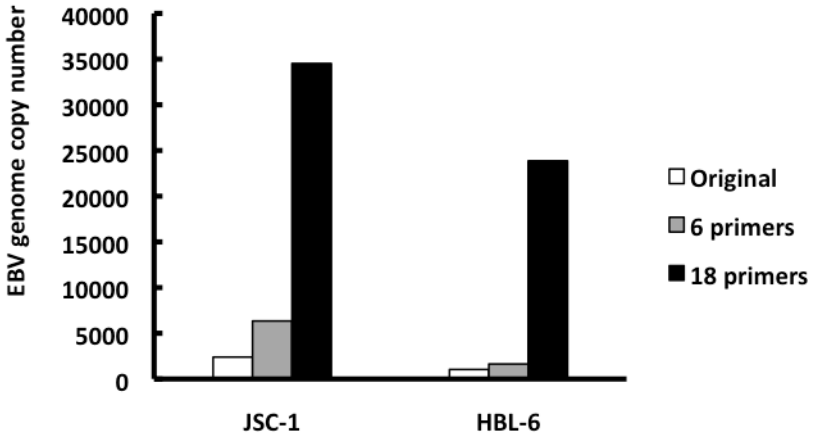


Figure 4.4. Comparison of the EBV genome copy number before and after Phi29 PCR. DNA from the Gardella gel slices was eluted, and the original copy number was determined by qPCR. Phi29 PCR was performed to increase the copy number to the concentration required for sequencing. Using 18 primers, the Phi29 PCR could increase the copy number of EBV by 10 fold.

4.2.5. Extracting EBV from PEL and BL cell lines by reactivation.

Other than separating the latent EBV episome from the cells by Gardella gel, EBV can also be amplified and enriched by reactivation of the lytic cycle. Chemical reagents such as VPA and BCR cross-linking were used to reactivate EBV into the lytic cycle. Viral DNA from the supernatant was extracted and concentrated. Two dually infected PELs, JSC-1 and HBL-6, and two BLs, Mutu I and Akata, were treated with VPA and BCR cross-linking respectively to reactivate EBV. By using VPA to reactivate EBV in the dually infected PELs, KSHV is also reactivated. EBV and KSHV copy number of these samples were determined by qPCR (Table 4.3). We calculated the relative amount of virus genome to host genome in base pairs. Despite the high copy numbers of EBV and/or KSHV, in comparison to a typical human genome (which is about 3Gbp), the total amount of virus DNA per sample was equivalent to 10^3 to 10^4 of human genome. This highlights the affect of even a small copy number of human genome will cause in the sequencing project. The human genome can dominate the total amount of DNA in the sample due to the sheer size, resulting in the majority of the NGS reads being non-EBV, decreasing the depth of sequencing and also increasing the difficulty of analysis. These EBV lytic stage samples were sequenced as multiplexed reaction with 6 per lane using a Genome

Analyser IIx (Illumina) at the WTSI. The samples were sequenced successfully and will be discussed in chapter 5 in detail.

Table 4.3. Copy number of EBV and KSHV from PELs and BLs

Cell line	Sample no.	EBV copy no./µl	KSHV copy no./µl	Total volume	Total EBV Genome equivalent (Gbp)	Total KSHV Genome equivalent (Gbp)
JSC-1	1	$2.9x10^6$	$1.7x10^7$	20µl	$1.0x10^4$	$5.8x10^4$
	2	$2.2x10^6$	$1.2x10^7$	20µl	$7.7x10^3$	$4.1x10^4$
	3	$2x10^6$	$3x10^6$	40µl	$1.4x10^4$	$2.0x10^4$
	4	$1.7x10^6$	$1.1x10^6$	40µl	$1.2x10^4$	$7.5x10^3$
HBL-6	1	$1.4x10^6$	$8.5x10^6$	20µl	$4.8x10^3$	$2.9x10^4$
	2	$9x10^5$	$6.5x10^6$	20µl	$3.1x10^3$	$2.2x10^4$
	3	$1x10^6$	$9.6x10^5$	40µl	$6.9x10^3$	$6.5x10^3$
	4	$1.9x10^6$	$8x10^5$	40µl	$1.3x10^4$	$5.4x10^3$
Akata	1	$1.8x10^6$	N/A	20µl	$6.2x10^3$	N/A
	2	$2.9x10^6$	N/A	20µl	$1.0x10^4$	N/A
	3	$6x10^6$	N/A	40µl	$2.1x10^4$	N/A
Daudi	1	$1.1x10^7$	N/A	20µl	$3.8x10^4$	N/A
	2	$1.5x10^7$	N/A	20µl	$5.5x10^4$	N/A
	3	$1.1x10^7$	N/A	40µl	$3.8x10^4$	N/A

4.2.6. Using Agilent SureSelect enrichment system

As mentioned previously, enriching or separating viral genomes from host DNA can result in a higher success rate of sequencing using Illumina. Earlier sections of this chapter have described separating the viral genome from the host DNA. However, the copy number of viral genome remained low despite using Phi29 DNA polymerase to enrich the separated viral DNA from each cell line. In addition, Phi29 DNA polymerase enrichment can only be used efficiently

on episomal samples. To sequence EBV from clinical samples, another method would therefore be required.

The Target Enrichment system (referred to as SureSelect) (Tewhey *et al., 2009)* developed by Agilent is currently used to sequence all human exons. Figure 4.5 illustrates the process of the SureSelect process. By adapting the procedure, this process was used to enrich and separate the EBV genome from other DNA in the sample. The SureSelect procedure can be divided into three stages. Firstly, the sample is sheared to desirable length, which is between 200-300bp long in this study, and mixed with target specific biotinylated “baits” (further detail in chapter 4.2.6), allowing the target DNA to hybridise with the baits. Secondly, the non-target DNA is removed from the sample by repetitive washing. And finally, the target DNA is enriched using a minimal number of PCR steps. This enrichment process removes the majority of the non-target DNA, and through minimal rounds of PCR, increases the target DNA further. The samples after the enrichment process can be sequenced as multiplexed reactions with 6 per lane using a Genome Analyser IIx (Illumina, Inc) at the WTSI.

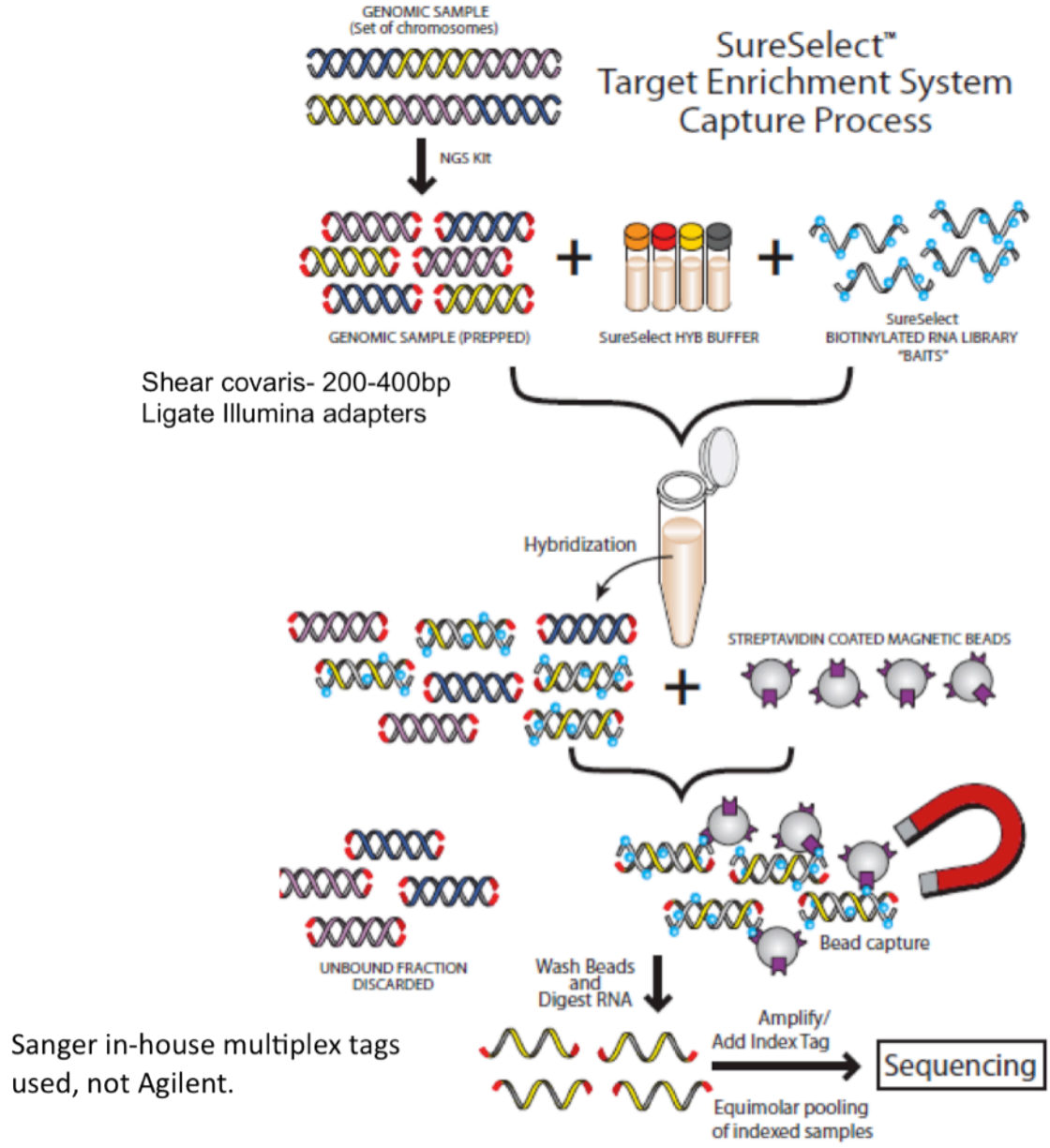


Figure 4.5. Flow diagram illustrating the Agilent SureSelect target enrichment process. DNA samples were sheared to 200-400bp length using the covaris, and ligated with Illumnina adapters, which then hybridised with the biotinylated baits, which was designed against the EBV type 1 and 2 reference genome from Genbank. The non-target DNA fragments were then removed by washing, post hybridisation by using streptavidin coated magnetic beads. Minimal PCR has then been carried out to amplify the DNA before sequencing the samples (adapted from Agilent SureSeclect website http://www.halogenomics.com/sureselect).

4.2.7. Bait design

One important feature of the SureSelect enrichment method is that there must be a reference sequence to enable the “bait” design. Currently there are five EBV genome sequences in Genbank, with two reference sequences

representing both type 1 and type 2 EBV genomes. Two sets of “baits”, one based on EBV type 1 (NC_007605) and another on EBV type 2 genome (NC_009334) were designed using Agilent eArray XD software. The baits were designed to be 120mer RNA probes, tiling five times across the whole genome. Each base in the genome is covered by five different probes, with a 24bp overlap between each probe. A total of 7154 and 7193 baits were designed and synthesised for EBV type 1 and type 2 respectively. Each library contains four sets of baits; two sets of EBV type 1 and type 2 each, which allows enrichment of either type 1 or 2 EBV genomes from the samples without prior determination of the virus type.

4.2.8. Whole Genome Amplification (WGA) before the SureSelect procedure

One disadvantage of using the SureSelect target enrichment methodology is that large quantities (3µg) of DNA template are required. For many clinical samples, the total amount of DNA available is less than 3µg. Phi29 DNA polymerase with primers specifically designed for EBV was used previously, to specifically enrich the copy number of EBV viral genomes. However, many clinical samples are stored as DNA, which has been sheared during the extraction process, therefore EBV specific primers cannot be used. Here, random primers (Genomiphi) were used instead to allow us to carry out whole genome linear amplification (WGA), resulting in an increase in the total quantity of DNA within the sample regardless the type, before processing the samples through SureSelect. The details of sample process are illustrated in the flow chart in figure 4.9.

4.2.9. Analysis - Computational

Large numbers of pair-end reads can be generated from second generation sequencing. In this study, the EBV DNA, purified from the cell line samples, was sequenced on an Illumina GAIIx system, while the PTLD samples were sequenced on an Illumina HiSeq system. Using Illumina GAIIx, the number of paired-end reads (which are 2 x 76bp long reads) returned from these samples were between 4.5 to 7 x 10^6 (table 4.4); however, this is the number of total

reads which includes both EBV and non-relevant reads, of which the majority are human DNA.

Figure 4.6 illustrates the computational pipeline that was developed to obtain EBV genome consensus sequences. Firstly, duplicate paired end reads were removed from the data set since the duplicates may skew mapping and affect the calling of polymorphisms. The read pairs were subsequently quality controlled (QC) (discussed in more detail later in this chapter) using QUASR v7 (figure 4.7). For samples sequenced with Illumina HiSeq, the percentage of read pairs passing QC is between 50 to 80% whilst the percentages of read pairs passing QC generated by GAIIx is higher, at around 85% (table 4.4). The difference in passing QC could be due to the type of samples that have been deep sequenced by these two Illumina machines. The cell line samples were sequenced using the GAIIx and the PTLD samples were sequenced using HiSeq. The DNA of PTLD blood samples had been stored in -20^{o}C for a period of time and may therefore have been more degraded than the DNA from the cell line samples. After the QC step, the read pairs were mapped to both EBV type 1 and 2 reference sequences. For the samples that were HiSeq sequenced, due to the large number of read pairs after QC, only a subset of the read pairs, selected randomly, were to aligned against the reference genomes using Burrows-Wheeler Aligner (BWA) (Chou *et al.*, 2010). Using the percentage coverage across the genome on both type 1 and type 2 genomes, the type of EBV genome in the samples is readily determined. The EBNA2 region, which contains large differences between type 1 and type 2 EBV sequences and has been generated in both type 1 and 2 mapping, was used to confirm the EBV type. The read pairs were then assembled to the correct consensus sequence according to the type of EBV selected. The mapping of read pairs back to the reference sequences were carried out using the Burrows-Wheeler Aligner (BWA) (Chou *et al.*, 2010). The QC process, and generation of majority consensus sequences, was carried out using QUASR v7 (http://sourceforge.net/projects/quasr). Finally, single nucleotide polymorphisms (SNPs) were determined by aligning the majority consensus sequences back to the reference sequence, and using VarScan v2.2.7 (Koboldt et al., 2009) to determine the allelic frequencies.

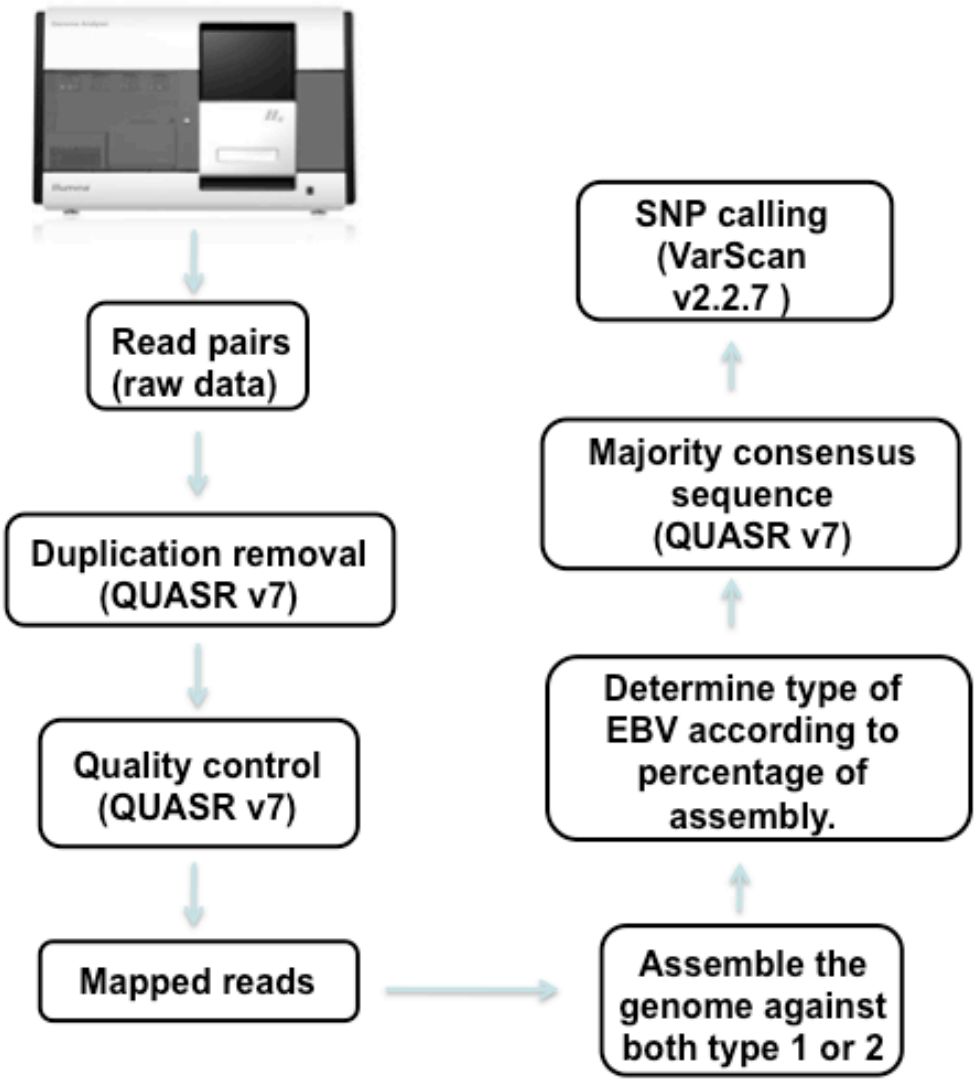


Figure 4.6. Flow chart of the computational analysis of read pairs generated by Illumina sequencing. Raw data were firstly processed through the quality control stage, which removes duplicate and trims low quality reads. Small percentages of the read-pairs were then assembled against both type 1 and type 2 EBV genomes. The type of EBV was determined depending on the percentage of mapping against these two genomes. All read-pairs were assembled to the correct EBV type and a majority consensus sequence obtained. The sequence was then aligned against the reference genome using Base-by-Base (Brodie *et al.*, 2004). All SNPs were identified using the Base-by-Base multi-genome comparison feature, and confirmed by QUASR 7, VarScan, and Sanger sequencing.

Figure 4.7 illustrates the process of QC. After removing the duplicate reads from the dataset, the dataset was divided into several smaller read sets from which the Illumina adaptor primers were removed. Read pairs generated from the Illumina carried a quality score (Phred score) on each individual base after the read pairs were generated. Phred scores were used to assess the quality of the sequence. Firstly, developed to assess the read quality in the Human Genome

Project, q Phred score is defined as a property that is logarithmically related to the base-calling probabilities. It works on the score system of 1-40 with exponential relationship to the probability. For example a Phred score 20 indicates a 1 in 100 probability of incorrect base call, and a Phred score 30 indicates 1 in 1000 probability of incorrect base call. Each base in a sequence had an associated Phred score, and the score of the total sequence was calculated according to all the Phred scores in the sequence. Phred score cut off during the assembly of consensus sequences was 30, indicating a 99.9% possibility of the reads and sequence being correctly called. The median of the Phred score of the last 10 bases of the read was calculated; if the median Phred score was below 30, the end base of the sequence was systematically removed and new median Phred score of the sequence recalculated. If the Phred score of the sequence reached 30 before the sequence became shorter than 50 base, the read was used in further assembly, otherwise the read and its pair were removed from the dataset.

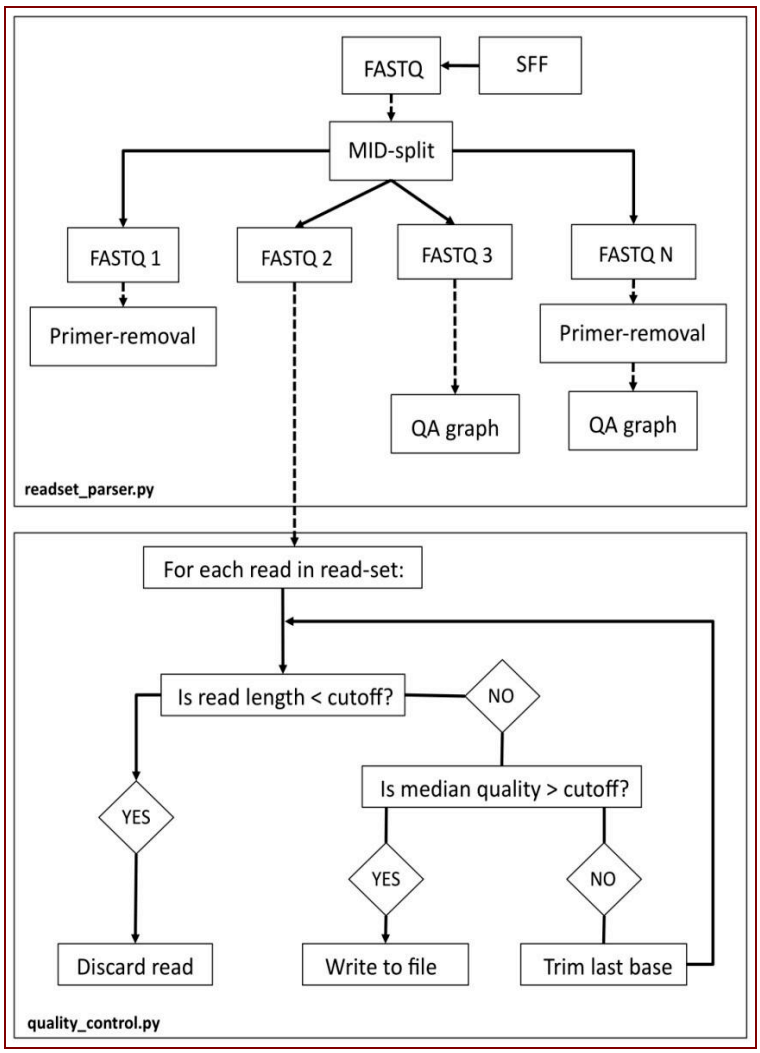


Figure 4.7. Flow chart illustrating the details of Quality control step after the read pairs had been generated from Illumina. The Phred score was set at 30, which means a 99.9% probability that the base is called correctly. The read length cut off point was set at 50 bases. The median Phred score of the last 10 bases of a read was used to determine the quality of the read, if it was below the cut off (30), the last base of the read was trimmed off and the median recalculated. This process was repeated until the median of the Phred score exceed the cut off, at this point if the trimmed read length was below the specific cut off (50 bases in this investigation), the read and its pair were discarded. Otherwise it was retained for mapping.

Once the majority consensus sequence was generated, it was aligned against the reference sequence according to the type. The location of single nucleotide polymorphisms (SNPs) and the polymorphic loci were called after the alignment using Base-by-Base software (Brodie et al., 2004). The SNPs were further confirmed by using VarScan v2.2.7 and QUASR v7 software (Koboldt et al., 2009) (http://sourceforge.net/projects/quasr) by investigating further into the depth and coverage of the read pairs of that particular base. Further analysis of the SNPs and polymorphisms will be discussed in chapter 5. SureSelect and

the computational analysis have been carried out in previous studies by Depledge et al., (2011).

4.2.10. Confirmation of genome enrichment

A pipeline for sequencing EBV whole genomes from both clinical and culture samples containing small amounts of EBV DNA has been established in this chapter. The samples which contained less than 3 µg of total DNA were whole genome amplified using Phi29 DNA polymerase and random primers, which then underwent the SureSelect process to enrich the target genome before deep sequencing. One of the sample types processed through this pipeline was the clinical blood samples of patients with post-transplant lymphoproliferate disease (PTLD). The detailed analysisof this data will be discussed in Chapter 5. We carried out qPCR on EBV and human GAPDH genome at various stages during this pipeline to investigate the degree of enrichment. The viral and human genome content was determined at five stages: the original sample after purification, after whole genome amplification (WGA), during the SureSelect process, pre hybridisation and also post hybridisation before PCR. Figure 4.8 shows the ratio between EBV and human DNA. Surprisingly, after WGA the ratio has lowered; indicating that WGA is apparently biased to amplify the human genome. The pre-hybridisation stage shows the same ratio as the WGA stage. Post hybridisation, three samples have an EBV to human DNA ratio increased to a maximum at seven logs. The last sample, despite not having as high increase in the ratio, also showed a log two increases. Depledge *et al*., (2011) also show similar results in the enrichment of the VZV samples. This indicates that the SureSelect process is successful in enriching the EBV substantially in the samples. However, WGA could be removed from the pipeline, as its effect to be seems detrimental.

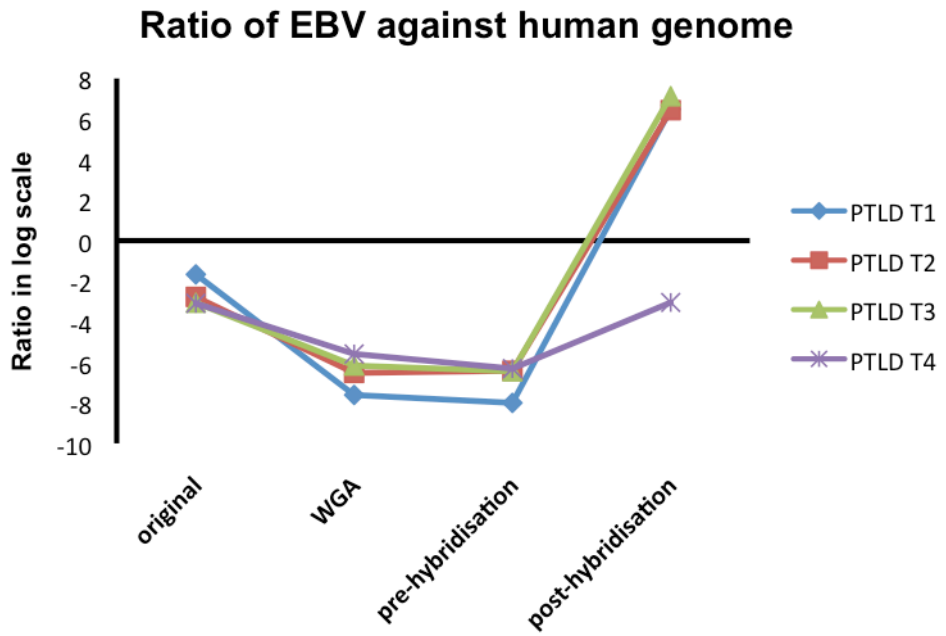


Figure 4.8. The ratio between EBV and human genome DNA during various stages of preparation prior deep sequencing. The blood samples are from the same patient, taken from four different time points. Original samples indicating the DNA sample, WGA after whole genome amplification by using Genomiphi. Pre and post-hybridisations are two stages in the SureSelect enrichment protocol, before the final PCR stage.

4.2.11. Illumina sequencing results

As discussed in chapter 4.2.8, Next Generation Sequencing such as Illumina generates a large amount of data. Two types of sequencing machines were used in this study, the Illumina GAIIx was used for all cell line samples, and Illumina HiSeq was used for the clinical PTLD data that had also been processed through the SureSelect process. Figure 4.9 illustrate the process by the type of sample and all the sample preparation procedures before the sequencing machines were used. A higher percentage of reads had passed QC from cell line samples (>70%) compared to the clinical samples (between 55-60%), for reasons that are unclear. With all the samples, over 163,000 bases of the EBV genome are covered at least once and at 10 times coverage, still over 160,000 bases of EBV are covered except for the PTLD T1 sample (where at 10x and 1x coverage is of 90% of the genome). The average depth of all samples, i.e. the number of times a base is sequenced on independent reads ranges between 87 to 1,000x for the clinical samples, to over 2,000x for the cell line samples. This is possibly due to the quality of the DNA in the different types

of samples. This is much higher than was seen using Sanger capillary sequencing (depth = 1), or in other next generation Illumina sequencing by Liu *et al.,* (2011) (average depth= 17). High depth coverage allows more accurate calling of SNPs, and also allowed us to study the evolution of EBV in more detail (further discussion in Chapter 5).

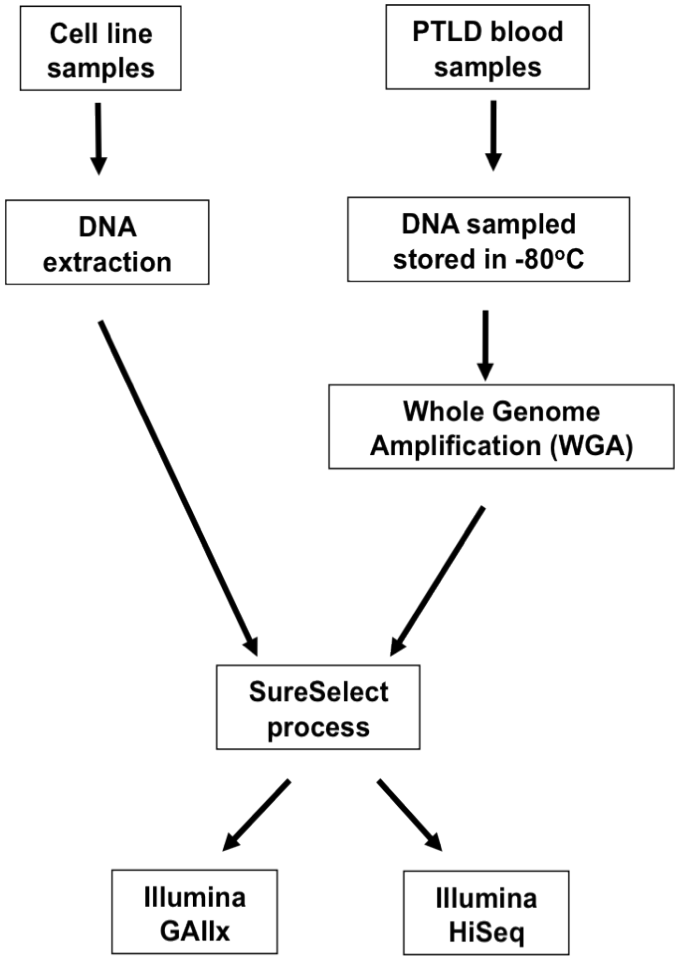


Figure 4.9. Flow chart illustrates the sample preparations before sequencing. Both cell line samples (lysate and supernatants) have DNA extracted using the Qiagen DNA extraction kit prior to processing through the Sureselect process. The samples were sequenced on the Illumina GAIIX machine. The PTLD blood samples were obtained as DNA samples which had been extracted previously. The DNA samples were processed with whole genome amplification and followed by the Sureselect process prior to sequencing on the Illumina HiSeq machine.

Table 4.4. Statistics of number of read-pairs post Illumina sequencing.

Sample	Sequencer	Total reads	Post-duplicate removal	Post QC	% reads post QC	EBV reads only				
						Uniquely mapped reads	Mean Read Depth	Coverage (1x)	Coverage (10x)	Coverage (100x)
PTLD T1	Illumina HiSEQ	150000	97731	88059	59%	76434	87	163225	152326	54225
PTLD T2	Illumina HiSEQ	393500	257000	229878	58%	198267	224	165588	161856	120638
PTLD T3	Illumina HiSEQ	845500	542857	491461	58%	418711	469	167260	162966	144381
PTLD T4	Illumina HiSEQ	1941750	1185264	1067767	55%	901910	1008	167564	164884	157419
PTLD 2035	Illumina HiSEQ	377500	220057	209680	56%	154475	179	165417	159524	112533
JSC1 SN	Illumina GAIIx	6836067	5161388	4868370	71%	3052570	2623	171686	169154	162931
JSC1 LYS	Illumina GAIIx	6864461	5140591	4835122	70%	3526008	3040	171542	167686	158791
HBL6 SN	Illumina GAIIx	4751548	3978809	3765969	79%	2767361	2378	171689	167796	160894
HBL6 LYS	Illumina GAIIx	6747765	5134756	4746384	70%	2666926	2306	171053	167210	154362
Daudi SN	Illumina GAIIx	4351267	3507323	3301960	76%	2740415	2411	167739	164075	159841
Akata SN	Illumina GAIIx	4516373	3835862	3637931	81%	3144703	2721	171149	168433	160955

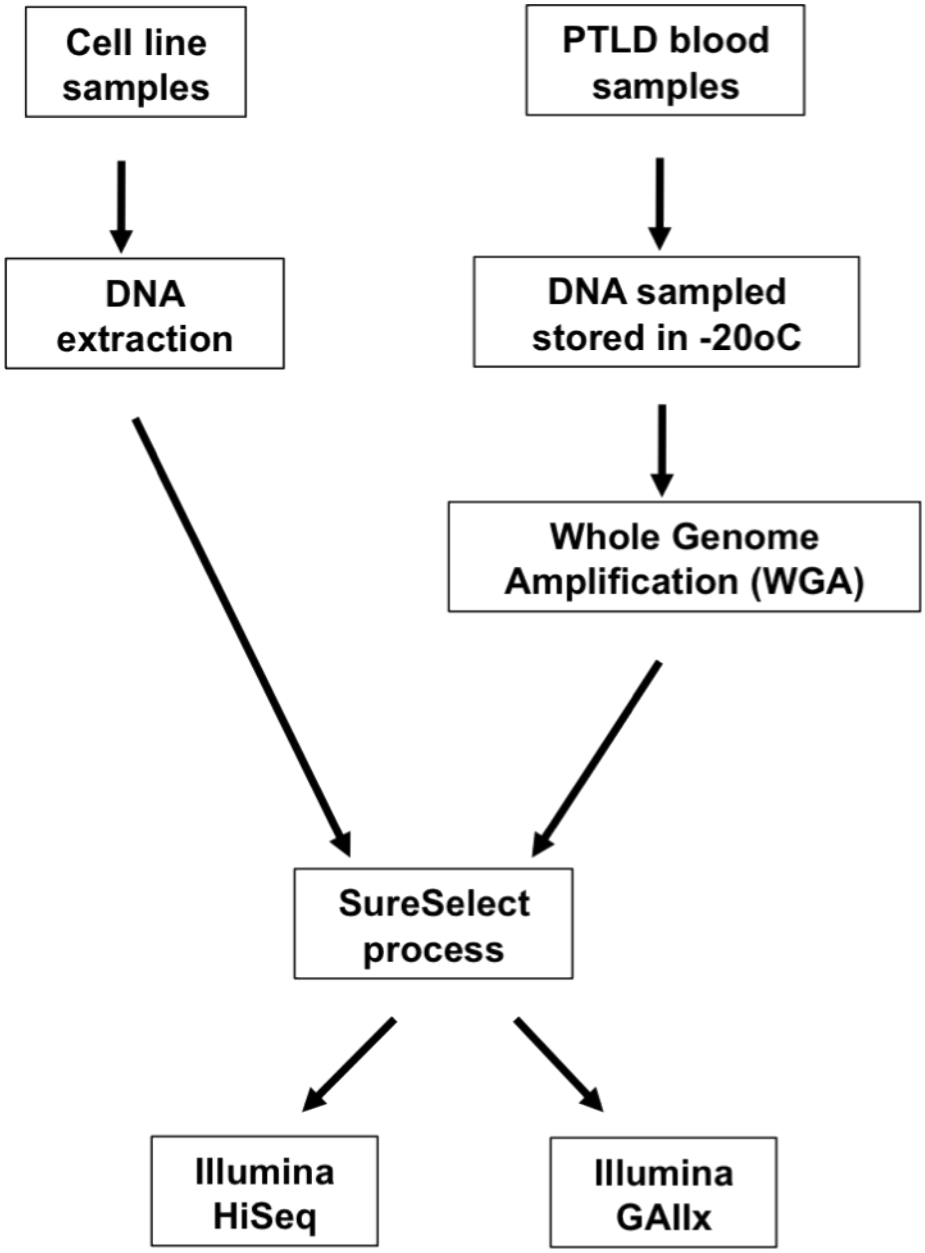


Figure 4.9. Flow chart of sample processing prior to next generation Illumina sequencing. The DNA of the cell line samples was extracted before the SureSelect process. The DNA from the PTLD blood samples was extracted and stored at -20^{o}C after the samples were taken. The DNA was whole genome amplified prior to the SureSelect process. After the SureSelect process, the cell line samples were sequenced using the Illumina GAIIx, with the PTLD samples sequenced using Illumina HiSeq.

4.3. Summary

Sequence variability of large DNA viruses has not been investigated in great detail due to difficulties in obtaining full genome sequences. Recent studies on HCMV (Renzette *et al.*, 2011) and VZV (Breuer *et al.*, 2010) have established that the mutation and polymorphisms occur more commonly in large DNA viruses than previously assumed. With five complete genomes in Genbank, it is difficult to determine the extent of polymorphisms in EBV. Previously the sequences of EBV have been generated by Sanger capillary sequencing (Sanger *et al.*, 1977b) and long PCR, which is time consuming and does not cover the genome in depth. A recent EBV genome GD2 has been generated using Next Generation Sequencing from clinical samples. However, with the host DNA intact in the sample, the result shows over 99% of the reads returned to be negative of EBV; therefore removing non-relevant DNA from the samples would reduce the time taken for, and cost of EBV sequencing and analysis.

In this chapter, several methods have been used to extract EBV for sequencing from various samples ranging from saliva to PBMCs. Both Phi29 DNA polymerase and the Agilent SureSelect target enrichment systems have been used to increase the copy number of the EBV. For sequencing samples containing large amounts of host DNA a pipeline has been established. The total DNA was increased by WGA, and target enriched by SureSelect prior to sequencing. However, after further investigation, it became apparent that WGA appear to be biased towards host DNA amplification and decreases the ratio of EBV to host DNA within the samples. WGA was therefore removed from our current pipeline.

Next Generation Sequencing generates large amounts of data. Careful computational procedures are required to sieve through the data and assemble the EBV genomes. Duplication and quality control process were carried out to reduce the possibility of sequencing error and remove low quality read pairs prior to assembly. Majority consensus sequences were generated using BWA software, mapping back to the reference genomes. *De novo* assembly was considered during the process of generating consensus sequencing; however, currently there is no ideal software for assembling EBV genome *de novo*. EBV contains many repetitive regions within the genome; it is difficult to assemble

the repetitive regions accurately with the BWA algorithm. Therefore the sequence of these repeated regions will require further investigation. Nevertheless, a complete assay for sequencing clinical samples has been set up in this chapter.

Chapter 5

Results: EBV Whole Genome Analysis

5.1. Introduction

DNA viruses are known to be less polymorphic than RNA viruses such as Human Immunodeficiency Virus (HIV) and the influenza viruses. HIV evolves at an extremely fast rate due to a high replication error rate and a large effective population size, together with high recombination level and fast replication rates (Mansky & Temin, 1995; Perelson *et al.*, 1996). The high mutation rates leads to many variants of HIV in a single infected individual in the course of one day (Robertson *et al.*, 1995a). Many of the polymorphisms and mutations have been linked to different selective pressures on HIV, for example immune evasion, disease progression and development of drug resistance.

Genetic variation in DNA viruses has not been studied extensively. Two herpesviruses have recently shown higher than expected genome variation in humans. Human cytomegalovirus (HCMV) is a member of the β-herpesviruses and, with a 236 Kb genome, is the largest human herpesvirus (Dolan *et al.*, 2004). A recent study suggested that every open reading frame (ORF) of HCMV showed levels of intra-host diversity, and quantify this variability by measuring nucleotide diversity shows the intra-host diversity of HCMV is comparable to RNA viruses such as HIV, with the frequency of variants in coding region sequence (CDS) being lower than in non-CDS. Although the source of diversity was unclear, the variation of HCMV also indicates selection within the population. Evidence of both positive and negative selection was determined suggesting positive selection contributes to the divergence across HCMV species, and negative selections occurs to maintain the genetic stability of the viral species (Renzette *et al.*, 2011).

Varicella-zoster virus (VZV) is a human α-herpesvirus which causes chickenpox (primary infection) and Shingles (reactivation). Of all the herpesviruses, VZV is the one with the highest number of whole genome sequences in GenBank. Excluding the variable regions, there are between 30-200 single nucleotide polymorphims (SNPs) between any two genomes, indicating a SNP every 4,000

to 600 bases. Around 30% of these SNPs result in amino acid changes, with uneven distribution across the genome (Tyler *et al.*, 2007).

Numerous studies have investigated the variation in EBV genes, the association with the geographical distribution and the affect on malignancy (See Chapter 1.5). Since there are limited numbers of EBV whole genome sequences, little is known regarding the variation and polymorphisms of EBV. In Chapter 4, we have discussed the method and the pipeline that were used to conduct Next Generation Sequencing (NGS) of EBV, resulting in whole EBV genomes being assembled from cell lines and clinical samples. In this chapter we concentrate on the investigation of the variation of the EBV whole genome majority consensus sequences (MCS), especially focusing on variant EBV type 1 sequences.

5.2. Results

5.2.1. EBV whole genome sequences

Following on from the computational analysis discussed at Chapter 4, a total of 11 EBV whole genome majority consensus sequence (MCS) were assembled. These include two BL cell lines: Akata and Daudi and two PEL cell lines: JSC1 and HBL6, with EBV from either the lytic cycle or latency. The other five sequences are from the whole blood samples of two Post Transplant Lymphoproliferative disorder (PTLD) patients. PTLD T1, 2, 3 and 4 are sequenced from the same patient at four different time points (table 4.4) during the development of the disorder with an increasing amount of viral load. These four samples were collected over a two months period, with the T3 and T4 samples taken 4 days apart (table 5.1).

Table 5.1. Time and viral load of the PTLD blood samples taken from the same patient

Sample number	Sample type	Date sample taken	Viral load (copies/ml)
PTLD T1	Blood	18.04.2008	362,000
PTLD T2	Blood	02.05.2008	2,352,720
PTLD T3	Blood	02.06.2008	6,300,000
PTLD T4	Blood	06.06.2008	13,100,000

All of the samples other than HBL6 and PTLD 2035 were assembled to the EBV type 1 reference sequence; HBL6 and PTLD 2035 were assembled to the EBV type 2 reference genome. This is dependent on the percentage of read pairs matching to either the type 1 or 2 reference sequences (details discussed in Chapter 4). Additionally, five EBV whole genome sequences from the saliva samples (Saliva 5, 5G, 7 and 7G) (Chapter 4), and another PEL sample BC1, were assembled at the WTSI using the same method. All but the BC1 sequences were assembled to the EBV type 1 genome. These sequences will not be discussed in detail, except in the phylogenetic studies later on in this chapter.

Both PTLD T1-4 and PTLD 2035 are sequenced from clinical blood samples.

The PTLD 2035 blood samples originated from a two-years-old male PTLD patient, and the PTLD T1-4 samples were from a one-year-old male PTLD patient at the first time of sampling. Both patients underwent renal transplants after renal failure, with the PTLD 2035 patient forming a tumour at the gut region and the PTLD T1-4 patient forming a tumour at the lung regions. It is unclear if either of the patients was sero-positive with EBV prior to the organ transplant.

5.2.2. Single Nucleotide Polymorphisms (SNPs) of the EBV genome

The EBV genome contains large numbers of nucleotide repeat sequences within the genome. Sequencing over these repeated regions (RDs) is difficult with both Sanger capillary sequencing (SCS) and Next Generation Sequencing (NGS). During genome assembly, it was apparent that the coverage across the RDs is low, with in some cases no reads mapping across the region. Previous studies have shown cell line type specific variation in the number of repeats in the EBV genome, however, the number of repeats cannot be correctly identified in the MCS constructed after NGS. As the reads from Illumina were aligned against an EBV reference sequence; the number of repeats in the newly assembled sequence will therefore be identical to the reference genome used. Therefore the variations that occurred over RDs are ignored as unreliable to ensure a more accurate understanding of the rest of the EBV genome.

Single nucleotide polymorphisms (SNPs) are called by comparing the majority consensus sequence against the relevant reference genome. Since the PTLD T1-T4 samples are from the same source over a short period of time, the SNP analysis is carried by comparing all four EBV sequences with the type 1 reference genome, therefore the number of SNPs of PTLD here refer to any variation that occurred in any of the four PTLD sequences combined. Table 5.2 shows the summary of the number of SNPs across the type 1 EBV assembled genomes, with the SNPs over RDs being discounted. Akata contains the highest number of SNPs in these samples, and JSC1 the fewest. All four samples contain higher number of SNPs within the CDS than the non-CDS; with more synonymous SNPs than non-synonymous SNPs (NS-SNPs). Synonymous change results in changes of nucleotides but does not result in

amino acid change; whereas non-synonymous change results in changes of both nucleotide and amino acid. The number of total SNPs from these samples indicates a SNP occurs between every 170-360bp in EBV genome, a much higher rate than seen in VZV suggesting EBV is more variable than previously expected. Between 26-30% of SNPs in CDS result in amino acid change.

Table 5.2. Number of Single Nucleotide Polymorphisms (SNPs) in coding (CDS) and non-CDS region of EBV genome of different disease background.

Type of SNPs	PTLD T1-4	JSC1	Akata	Daudi
CDS Synonymous	412	284	638	471
CDS Non-synonymous (amino acid change)	161	124	245	169
Non-CDS	273	237	358	245
Total number	683	521	996	716

5.2.3. Latent and Lytic EBV genome from PELs

Both the lytic virus genome and the latent virus genome of EBV were sequenced from JSC1 and HBL6 cell lines by sequencing the EBV genome from supernatant post stimulating EBV reactivation (lytic) or directly from the lysate of the latent cells (latent) (see Chapter 4 for detail). The majority consensus sequences of these two types of EBV genome were aligned using Base-by-Base (Brodie *et al.*, 2004). Following removal of RDs, there was no difference between the latent and lytic EBV genomes in either JSC1 or HBL6 cell lines. This indicates our method is robust and that there is no observable genetic variation in EBV genomes at the consensus level during lytic reactivation. The latent virus genomes of EBV from the BL cell lines Akata and Mutu were not sequenced, therefore the BL cell lines were not included in this analysis.

5.2.4. Non-Synonymous SNPs of cell lines distribute unevenly across the genome

The hundreds of non-synonymous (NS)-SNPs in the EBV type 1 sequences show uneven distribution across the genome. Comparing the distribution of NS-SNPs of the cell line sequences, it is clear that the distributions show a similar pattern between the three cell lines, especially between Daudi and JSC1 EBV sequences (figure 5.1). The amount of difference between the genomes suggests the reference genome is not representative, with EBV from JSC1 and Daudi revealing the unrepresentative regions. The Akata EBV genome showed a different NS-SNPs distribution. JSC1 and Daudi EBV sequences are more similar to the type 1 reference genome than the Akata EBV sequence. From the pattern of NS-SNPs distribution, it is clear that regions of the genome contain either higher numbers of NS-SNPs or no NS-SNPs, suggesting there are regions that are more variable or conserved in the genome (figure 5.2).

The EBV genomes of Daudi and JSC1 cell lines show no or low numbers of NS-SNPs in certain regions across the genome, such as regions between BHRF1 and BFRF1 genes, BLLF2 and BZLF1 genes, and BGLF4 and BGLF2 genes. The EBNA 3A, B, C and BZLF1 genes in the two genomes are identical to the reference gene. The majority of the highly polymorphic genes are located in the first 50% of the EBV genes (starting from the TR region at the 5' end of the genome and excluding the W repeat region), such as at the BNRF1, EBNA2, BPLF1, BOLF1, BRRF2, EBNA1 and BBLF1 genes. The BLLF1 and BSLF1 genes are also highly polymorphic in the EBV genome of Daudi origin, but not in that of JSC1 origin. EBNA2 is the major protein for determining the type of EBV, and differs by 55% in the protein sequences of the two EBV types. Here we show that within the same type of EBV genome the sequence of EBNA2 gene also varies. The EBNA1 gene is known to be highly polymorphic in the EBV genome, and has been studied extensively previously. Here we confirmed the highly variable nature of EBNA1 and showed more genes of the EBV genome that are highly polymorphic. It is unclear what selective pressures resulted in the conserved and variable genes and regions in the EBV genome or the functional significance of these polymorphisms.

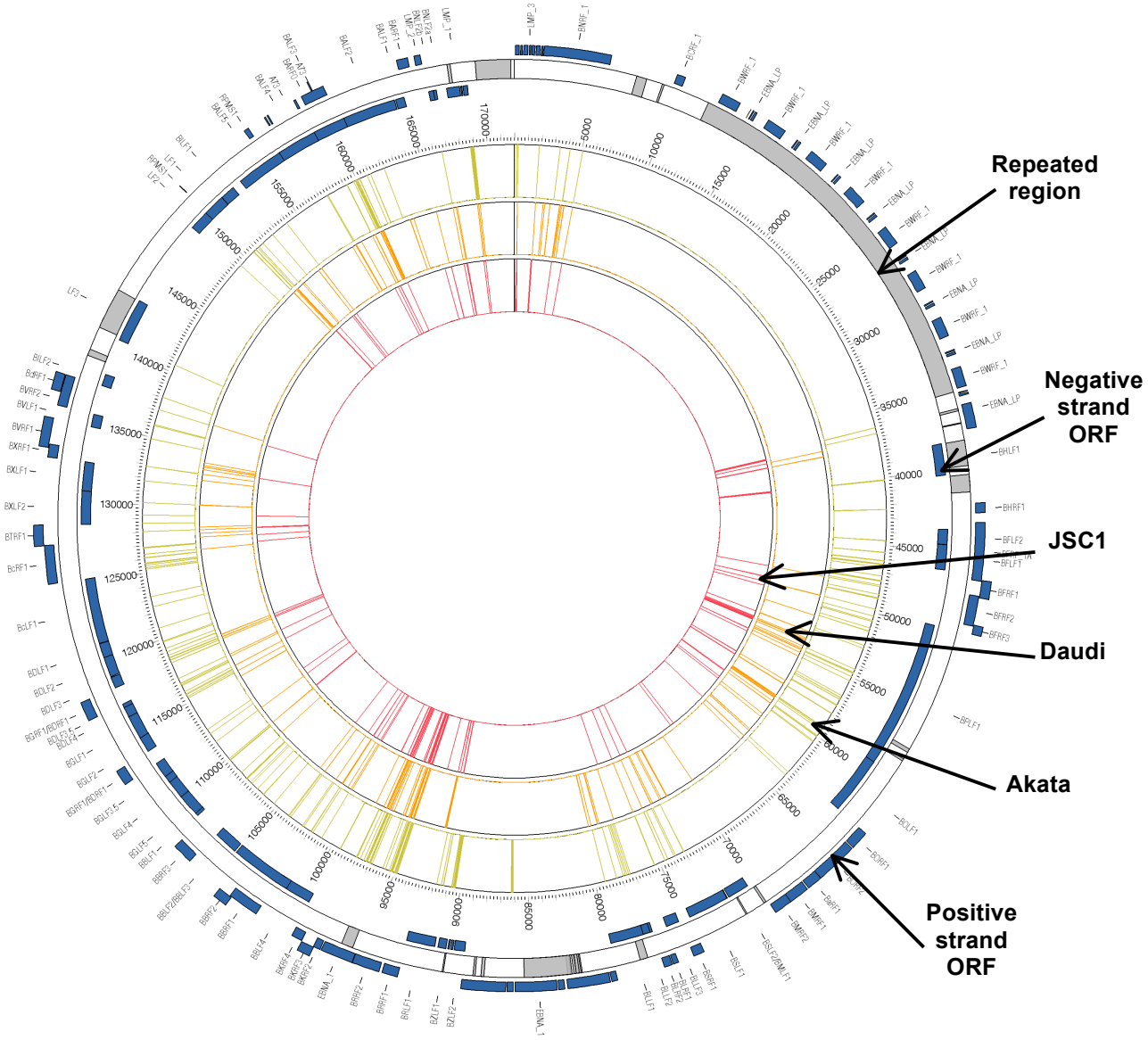


Figure 5.1. Circos plot showing polymorphisms of EBV in PEL (JSC1,) and BL (Akata and Daudi) cell lines in comparison to the EBV type 1 reference genome. The outermost ring represents the reference genome, with the three inner rings representing Akata (yellow), Daudi (orange) and JSC1 (red) respectively. The lines in each of the three rings represent non-synonymous (NS)-SNPs in the coding region, with the NS-SNPs located at the repeat regions removed prior to this analysis. The distribution of NS-SNPs shows a similar pattern in the EBV genomes of JSC1 and Daudi. There are clearly more NS-SNPs in the EBV genome of Akata, which shows a wider distribution of NS-SNPs across the genome.

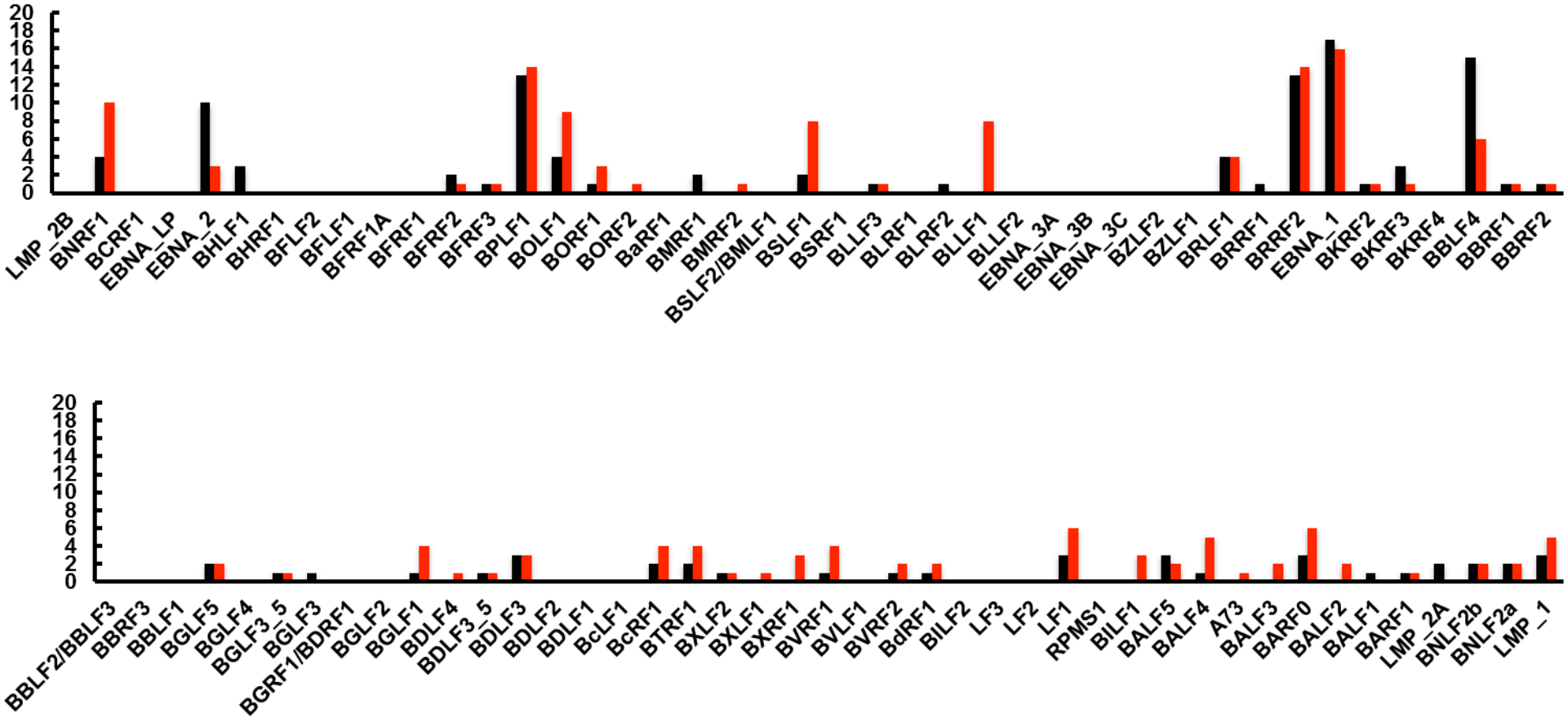


Figure 5.2. Number and distribution of non-synonymous SNPs of EBV type 1 genomes from cell line samples: JSC1 (black) and Daudi (red), compared to the EBV type 1 reference genome. The NS-SNPs occurring at the repeat regions have been discounted. The genes are listed in order on the EBV type 1 reference genome with the number of NS-SNPs on the Y-axis. The general distribution of NS-SNPs in these two cell lines are similar across the genome. Both the EBV of JSC1 and Daudi show regions of low variation (low or no NS-SNPs in the region), such as between BHLF1 and BFRF1 and between BLLF2 and BRLF1 (containing all the EBNA3 genes).

5.2.5. Non-Synonymous SNPs distribution in PTLD samples

We wished to determine if EBV genome variation could be observed over time in an infected individual. The study of the samples from the same PTLD case over four time points by full EBV genome sequencing (PTLD T1-4) is the first longitudinal EBV whole genome study. Over the two months period of sampling, a total of 682 SNPs occurred through the four consensus whole genomes with 161 NS-SNPs resulting in amino acid changes when compared against the type 1 reference genome. These NS-SNPs either occured only at one time point or at multiple time points. As with analysis of the EBV cell line sequences, the NS-SNPs were not distributed evenly across the genome, but again show more conserved and variable regions of the genome (figure 5.3). The majority of the NS-SNPs occur in the first half of the genes starting from the 5' end, excluding the RDs. In the EBV genome, BNRF1, EBNA2, BFRF2, BPLF1, BOLF2, BLLF1, BRRF2, EBNA1 and LMP1 were highly polymorphic, showing a similar pattern as the EBV genomes from cell lines. The distribution of the NS-SNPs in EBV sequences from the PTLD samples also shows highly conserved regions with little or no NS-SNPs. The distribution of S-SNPs follows similar pattern as the NS-SNPs; with few exceptions, such as BcLF1, which contains a high number of S-SNPs but very low number of NS-SNPs.

Table 5.3 shows the number of SNPs occurring in each gene of the EBV genome of the PTLD EBV sequence. The SNPs occurring in the repeated regions have been removed prior to this analysis. The majority of the NS-SNPs occur at the genes encoding tegument proteins and glycoproteins. The tegument proteins and glycoproteins also contain a higher number of NS-SNPs (>five) (highlighted in orange), although EBNA1, 2, LMP1 and BFRF2 genes also contain >five NS-SNPs. Interestingly, EBNA2, which is the most highly variable gene between the EBV type 1 and 2 genomes, only contains 9 NS-SNPs. The most highly polymorphic gene of the PTLD EBV sequence is EBNA1. Despite the fact that EBNA1 contains a large repeat region within the gene; there are 17 NS-SNPs within the non-repeat region of the gene. Subsequently, the second most polymorphic gene, BRRF2, contains 14 NS-SNPs within the non-repeat region of the gene. The highly polymorphic phenotype of EBNA1 is the same as the EBV from cell lines and also has been observed previously (Bell *et al.*, 2008); although it is the first time that other

genes especially the tegument and glycoproteins, have been also shown to be highly polymorphic in the genome. All the EBV genomes that have been sequenced from either PTLD or cell line backgrounds have shown that the majority of the variation occurs in the first half of the EBV genome (first half of the EBV genes starting from terminal repeat (TR)), although the reason and evolutionary pressure behind this phenotype is unclear.

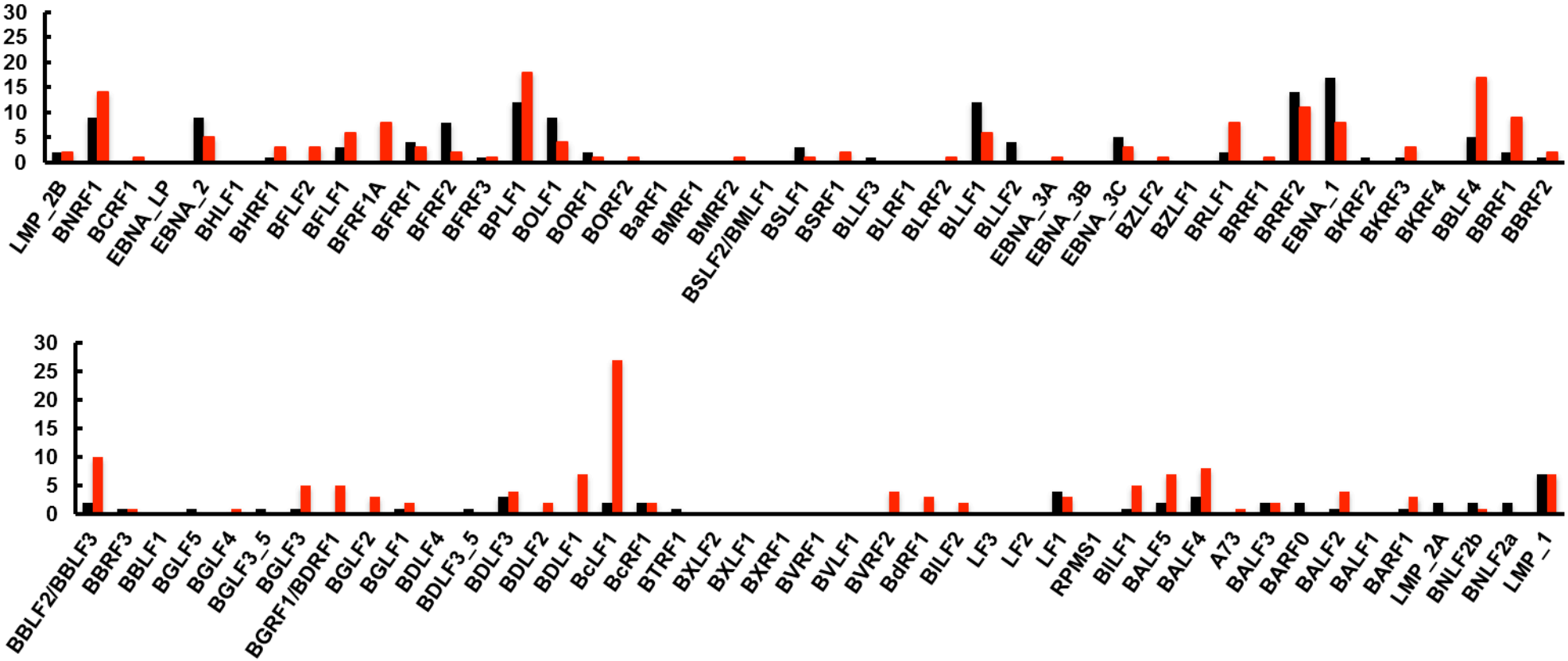


Figure 5.3. The NS-SNPs (black) and S-SNPs (red) distribution across the PTLD EBV genomes in comparison to the EBV type 1 reference sequence. The repeated regions have been removed from the genome prior the analysis. The number of NS-SNPs is on the Y-axis. The SNPs occurred at either a single time point multiple time points throughout the two month sampling period. This shows the uneven distribution of both NS and S-SNPs across the PTLD EBV genome. The genes are listed in order of the distance from the TR in the linear format. NS-SNPs occur more frequently in genes such as BNRF1, EBNA2, BFRF2, BPLF1, BOLF2, BLLF1, BRRF2 and EBNA1. The distribution of S-SNPs in the genome occur in a similar pattern to the NS-SNPs, with the exception of BcLF1 gene which encodes for major capsid protein, showing the highest number of S-SNPs. The distribution of SNPs also showed more variable (high number of SNPs) and conserved (low or no SNPs) regions of EBV genome.

Table 5.3. Details of the number of the NS-SNPs in each EBV gene of the PTLD whole genomes at either one or multiple time points throughout the 2 month sampling period.

Genbank name	Functions	Number of NS-SNPs	Number of S-SNPs
LMP_2B	Latent membrane protein 2A/B	2	2
BNRF1	Tegument protein	9	14
BCRF1_	Viral intreleukin (IL) 10	0	1
EBNA_LP	Glycoprotein L (Ubiquitin specific cystein protease)	0	0
EBNA_2	Glycoprotein L (Ubiquitin specific cystein protease)	9	5
BHLF1	Glycoprotein L (Ubiquitin specific cystein protease)	0	0
BHRF1	Bcl-2	1	3
BFLF2	Capsid maturation protein (AD169-UL53)	0	3
BFLF1	DNA packaging protein (AD169-UL52)	3	6
BFRF1A	DNA packaging protein (AD169-UL51)	0	8
BFRF1	Tegument protein (HHV4-BFRF1)	4	3
BFRF2	Capsid protein (AD169-UL49)	8	2
BFRF3	Capsid protein (C500-ORF65)	1	1
BPLF1	Tegument protein	12	18
BOLF1	Tegument protein (B95-8-BOLF1)	9	4
BORF1	Assembly/DNA maturation protein	2	1
BORF2	Ribonucleotide reductase large subunit	0	1
BaRF1 (UL40)	Ribonucleotide reductase small subunit	0	0
BMRF1	DNApol processivity subunit (GammaHerpes)	0	0
BMRF2	Unknown (B95-8-BMRF2)	0	1
BSLF2_BMLF1	Immediate-early phosphoprotein (B95-8-BMLF1)	0	0
BSLF1	Helicase/primase complex (AD169-UL70)	3	1
BSRF1	Unknown (AD169-UL71)	0	2
BLLF3	dUTPase	1	0
BLRF1	Envelope glycoprotein (AD169-UL73)	0	0
BLRF2	Unknown (C500-ORF52)	0	1
BLLF1	pp150 Tegument protein (AD169-UL32)	12	6

BLLF2	Unknown (B95-8-BLLF2)	4	0
EBNA_3A	HHV-4-B95_8-BLRF3	0	1
EBNA_3B	Glycoprotein L (Ubiquitin specific cystein protease)	0	0
EBNA_3C	Glycoprotein L (Ubiquitin specific cystein protease)	5	3
BZLF2	Glycoprotein gp-12	0	1
BZLF1	Immediate early protein	0	0
BRLF1	Replication and Transcription activator	2	8
BRRF1	Unknown (B95-8-BRRF1)	0	1
BRRF2	Glycoprotein L (Ubiquitin specific cystein protease)	14	11
EBNA_1	Episome maintenance protein	17	8
BKRF2	Glycoprotein L (GammaHV)	1	0
BKRF3	Uracil-DNA glycosylase	1	3
BKRF4	Immediate-Early Phase Viral Replication	0	0
BBLF4	Helicase/primase complex (AD169-UL105)	5	17
BBRF1	DNA packaging protein (AD169-UL104)	2	9
BBRF2	Unknown (AD169-UL103)	1	2
BBLF2_BBLF3	DNA helicase/primase complex (C500-ORF40)	2	10
BBRF3	Glycoprotein M	1	1
BBLF1	Myristylated virion protein (AD169-UL99)	0	0
BGLF5	DNase	1	0
BGLF4	Serine/threonine protein kinase	0	1
BGLF3_5	Unknown (C500-ORF35)	1	0
BGLF3	Unknown (AD169-UL95)	1	5
BGRF1_BDRF1	DNA packaging protein/terminase (AD169-UL89)	0	5
BGLF2	Tegument protein (AD169-UL94)	0	3
BGLF1	DNA packaging protein (B95-8-BGLF1)	1	2
BDLF4	Unknown (AD169-UL92)	0	0
BDLF3_5	Unknown (C500-ORF30)	1	0
BDLF3	pp150 Tegument protein (AD169-UL32)	3	4
BDLF2	Intercellular Viral Spread	0	2
BDLF1	Capsid triplex protein	0	7

BcLF1	Major capsid protein (AD169-UL86)	2	27
BcRF1	Unknown (AD169-UL87)	2	2
BTRF1	Unknown (B95-8-BTRF1)	1	0
BXLF2	Glycoprotein H	0	0
BXLF1	Thymidine kinase	0	0
BXRF1	Fusion protein (AD169-UL76)	0	0
BVRF1	DNA packaging protein (AD169-UL77)	0	0
BVLF1	Unknown (AD169-UL79)	0	0
BVRF2	Proteinase/capsid scaffold protein (AD169-UL80)	0	4
BdRF1	Proteinase/capsid scaffold protein (AD169-UL80)	0	3
BILF2	Unknown (B95-8-BILF2)	0	2
LF3	Glycoprotein L (Ubiquitin specific cystein protease)	0	0
LF2	Unknown (C500-ORF11)	0	0
LF1	Unknown (C500-ORF10)	4	3
RPMS1	HHV-4-WT-RPMS1	0	0
BILF1	Putative G-protein coupled receptor (C500-A5)	1	5
BALF5	DNA polymerase catalytic subunit	2	7
BALF4	Glycoprotein B	3	8
A73	HHV-4-WT-A73	0	1
BALF3	DNA packaging protein (AD169-UL56)	2	2
BARF0	Unknown (HHV4-WT-BARF0)	2	0
BALF2	ssDNA binding protein	1	4
BALF1	Unknown (B95-8-BALF1)	0	0
BARF1	Unknown (B95-8-BARF1)	1	3
LMP_2A	Latent membrane protein 2A/B	2	0
BNLF2b	Unknown (B95-8-BNLF2b)	2	1
BNLF2a	Unknown (B95-8-BNLF2a)	2	0
LMP_1	Glycoprotein L (Ubiquitin specific cystein protease)	7	7

5.2.6. EBV evolution in a PTLD patient

The NS and synonymous (S)-SNPs of the PTLD EBV genome sequences have show uneven distribution, with the EBNA1 gene being the most highly polymorphic gene in the PTLD EBV genome in comparison to the reference genome. As the four PTLD samples were taken over two months, we decided to investigate the variation these samples accumulated through time. Figure 5.4 shows both the NS-SNP (red line) and S-SNPs (black line) in CDS, that occurred compared with only the previous time point, not the reference genome. Despite the short time between samples, changes are observed throughout the two months. The largest amount of variation occurs between T1 and T2, despite the samples being taken only two weeks apart (table 5.1), with seven NS-SNPs occurring between these two sequences. The viral load of these two samples also showed the largest increase over the four time points. The lowest number of SNPs in CDS occurred between T3 and T4, with only one NS-SNP occuring. This is not surprising since the two samples were taken only four days apart. However, variability of EBV sequences over such short period of time is interesting. Throughout these four time points of PTLD sampling, the SNPs occured in similar regions of EBV genome, possibly indicating higher selective pressure on different genes.

Despite EBNA1 gene being the most polymorphic gene in EBV genome, when comparing EBV genomes of PTLD and cell lines with the reference genome; it is not the gene with the highest variation rate over time. When we investigated the changes of genes through time, EBNA1 remains relatively constant in the two months of PTLD sampling (figure 5.4), showing little variation through time when compared to the sequence at the previous time point. The majority of the NS-SNPs between two time points of sampling occurred in the BOLF1 gene, with six NS-SNPs between T1 and T2, three NS-SNPs between T2 and T3 and one NS-SNP between T3 and T3 (table 5.4). A total of nine NS-SNPs occur over time and are located in the BOLF1 gene. Notably, at location 61412, the nucleotide changed from C at T1, to A at T2 and back to C at T3, resulting in the codon changing from glutamic acid to a stop codon and back, suggesting that the mutation introduced at T2 may be selected against and therefore reverts back to the original nucleotide. BOLF1 encodes for a tegument protein with the function being unclear, although it is suspected that it is involved in

virus assembly and budding. BOLF1 also shares homology with the HLA-DQw8 beta chain (Sairenji *et al.*, 1991). The total number of NS-SNPs that occurred reduced through time, indicating the rate of variation, slowed towards the end of sampling period. However, since the study only consists of one patient over a two month period, it is difficult to be confident about the alteration of the genome variation rate over time of EBV genome.

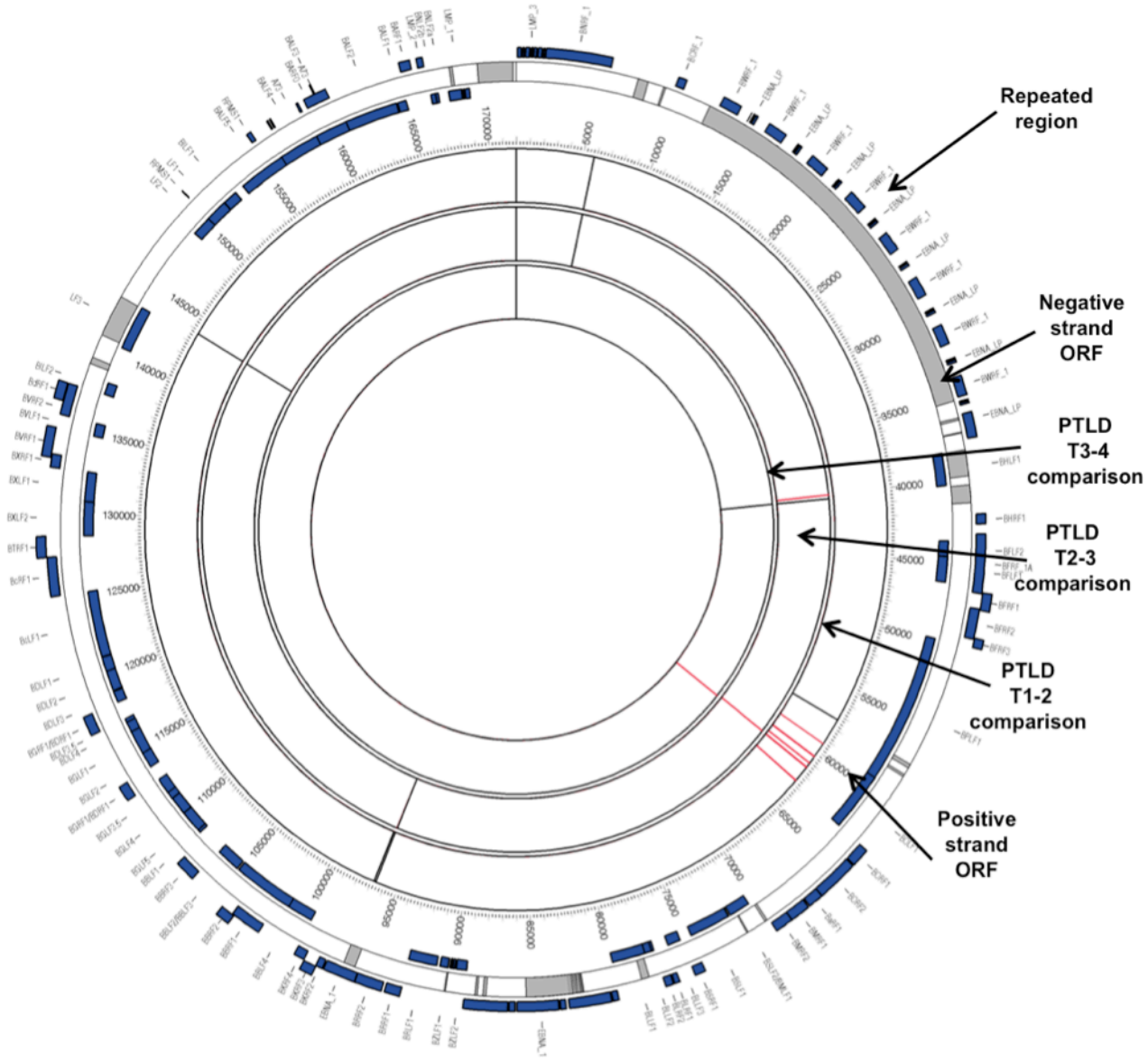


Figure 5.4. Circos plot representing the SNPs occurring between the time points for the PTLD patient. The red lines represent non-synonymous changes, and the black lines represent synonymous changes. The outer ring represents the reference type 1 EBV genome, and the second ring represents the SNPs between time point 1 (T1) and 2 (T2), the third ring represents the SNPs between time point 2 (T2) and 3 (T3). The most inner ring represents the SNPs between time point 3 (T3) and 4 (T4). The SNPs in the repeated region should be treated with caution since the coverage in these regions is low.

Table 5.4. The NS-SNPs occuring over time in the PTLD patient, showing the nucleotide change and the resulting amino acid change.

Location	Gene	PTLD T1		PTLD T2		PTLD T3		PTLD T4	
		Nucleotide	Amino acid	Nucleotide	Amino acid	Nucleotide	Amino acid	Nucleotide	Amino acid
59336	BOLF1	G	Aspartic acid	A	Tyrosine				
60452	BOLF1	G	Proline	A	Serine				
60484	BOLF1	C	Glycine	A	Valine				
61054	BOLF1	C	Glycine	A	Valine				
61412	BOLF1	C	Glutamic acid	A	Stop				
62653	BOLF1	A/G	Aspartic acid/ Alanine	G	Alanine				
40031	BHLF1			A	Leucine	C	Arginine		
61412	BOLF1			A	Stop	C	Glutamic acid		
96649	EBNA1			A	Serine	G	Glycine		
61573	BOLF1					T	Glutamic acid	C	Glycine

5.2.7. Polyclonal origin of EBV in blood samples of PTLD patients

Previous studies have shown that EBV in PTLD blood samples can be either poly- or oligoclonal in origin, with a very low probability of being monoclonal (Anderlini *et al.*, 2004). We have assembled the majority consensus sequence from the blood samples of PTLD patients where any nucleotide over 50% of the post QC reads is called as the majority nucleotide. SNP calling can only therefore show the change in majority base over time. This masks fluctuations in minority nucleotides which may also be informative. Minority viral variants have been shown to be important in RNA viruses, which have been associated with disease progression, treatment failure and viral evolution (Vignuzzi *et al.*, 2006; Zagordi *et al.*, 2010). However, the minority nucleotide variants in DNA viruses have not been extensively studied.

Therefore we determined the changes in nucleotide frequency over time at the NS-SNPs that occurred on the EBV genome sequence from PTLD samples. Here we define the occurance of minor nucleotide being if it exceeds 5% of total reads at any time point. Figure 5.5 shows six examples of the minority nucleotide changes over time for the EBV genome. A total of ten locations with fluctuations in percentage of nucleotide expression were observed at any time point. The six examples shown were chosen as they have reads covering the genome position at all four times. The appearance of minority nucleotides occurred in the EBNA2, BOLF1, LMP1 and EBNA1 genes. These four genes have been shown to be highly polymorphic when compared to the reference genome, with BOLF1 not only highly polymorphic but also having the highest mutation rate. Figure 5.5.A and B, both located in the EBNA2 gene, showed the abundance of minority species increasing over time. Figure 5.5.C, D, and E are located in the BOLF1 gene. Figure 5.5.C shows fluctuation of percentage of minority species over time, with Figure 5.5.D, E and F showing abundance of a minor species at only one point and then not at other three.

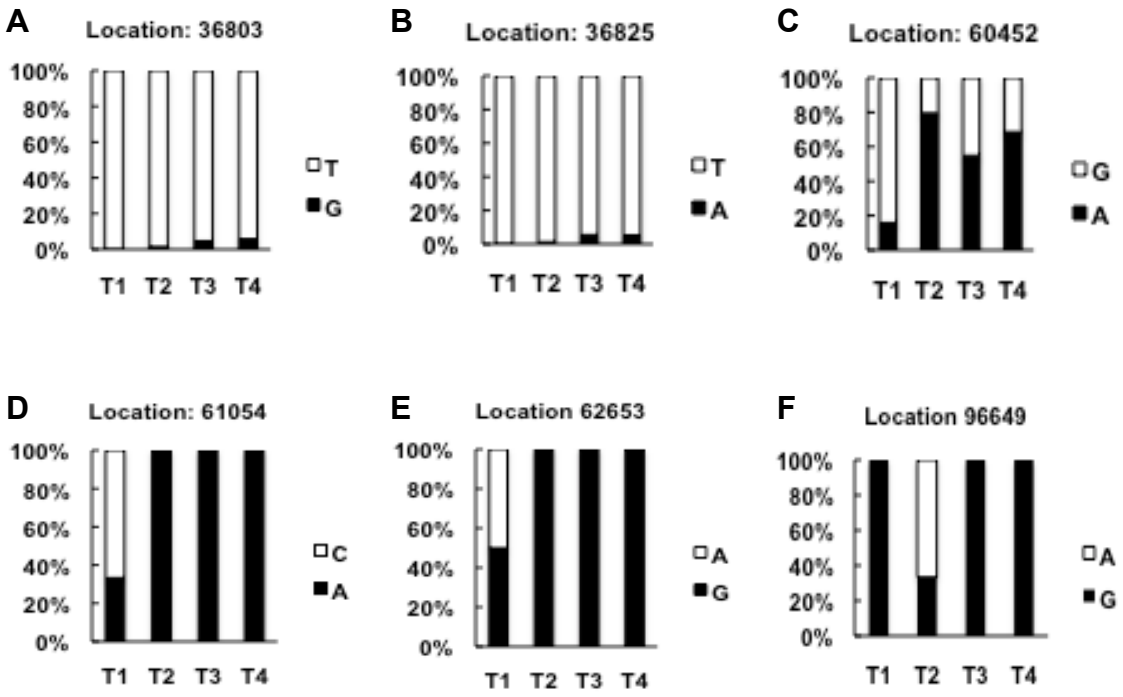


Figure 5.5. Minority species and the alteration of nucleotide abundance at different time points in the EBV PTLD genomes. Black bars represent the nucleotide in the reference sequence and white bars represent the alternative nucleotide in the samples sequenced. A base is called as a majority consensus sequence when the percentage of the nucleotide is over 50%. A) and B) both located in the EBNA2 gene, both show increase of nucleotide guanine (G) and adenine (A) expression respectively over time, but not sufficiently to alter the base calling. The large expression of thymine (T) in A) results in the amino acid changing from methionine to iosleucine, and in B) results in serine instead of Threonine. C), D) and E) are located in the BOLF1 gene, in C), the expression of G increased at T2, and caused the amino acid to change from serine to proline, and although the percentage of G flunctuate at T3 and T4, it remains over 50% of nucleotides in the sample. In D), expression of cytosine (C) at T1 caused the amino acid to change from valine to glycine, however, at T2,3 and 4, the expression of C cannot be detected. In E) a similar situation as D) occurred only the abundance of A is 50%, therefore there were 50% aspartic acid. At the later time point, the abundance of G returns to 100%, the amino acid remaining Alanine. F) is in the EBNA1 gene, with expression of A introduced at T2 and causing the amino acid change from glycine to arginine.

5.2.8. Phylogenetic analysis of whole EBV genomes and coding regions of the genome

With the EBV whole genome sequences assembled from different diseases backgrounds, it is interesting to see the phylogenetic relationship using whole genomes. A total of 11 majority consensus sequences were assembled from PEL, BL and PTLD samples, with the additional five assembled at WTSI, from PEL and saliva samples (table 5.5). These included both type 1 and type 2 EBV genomes. Combined with the EBV whole genome sequences available from GenBank: GD1 and GD2, both of NPC origin and two reference EBV genomes, there are total 20 EBV sequences for analysis (table 5.5). As previously described, the EBV majority consensus sequence was assembled against either EBV type 1 or type 2 reference genome from GenBank, which is determined on the percentage of the reads that have been successfully mapped back to the reference genome. All EBV from the saliva samples, JSC1, Akata, Daudi, PTLD T1-4 were assembled to the EBV type 1 reference genome, with GD1 and GD2 from GenBank also being type 1 EBV. HBL6, BC1 and PTLD 2035 EBV genomes have been assembled to the EBV type 2 reference genome.

Table 5.5. EBV samples in phylogenetic study. Genomes 1-16 were generated using NGS, with genomes 17-20 from Genbank.

	Name	EBV type	Disease associated
1	JSC1 SN	1	PEL
2	JSC1 Lysate	1	PEL
3	HBL6 SN	2	PEL
4	HBL6 Lysate	2	PEL
5	BC1	2	PEL
6	Akata	1	BL
7	Daudi	1	BL
8	PTLD 2035	2	PTLD
9	PTLD T1	1	PTLD
10	PTLD T2	1	PTLD
11	PTLD T3	1	PTLD
12	PTLD T4	1	PTLD
13	Saliva 5	1	N/A
14	Saliva 5G	1	N/A
15	Saliva 7	1	N/A
16	Saliva 7G	1	N/A
17	GD1	1	NPC
18	GD2	1	NPC
19	EBV reference 1	1	IM and BL
20	EBV reference 2	2	BL

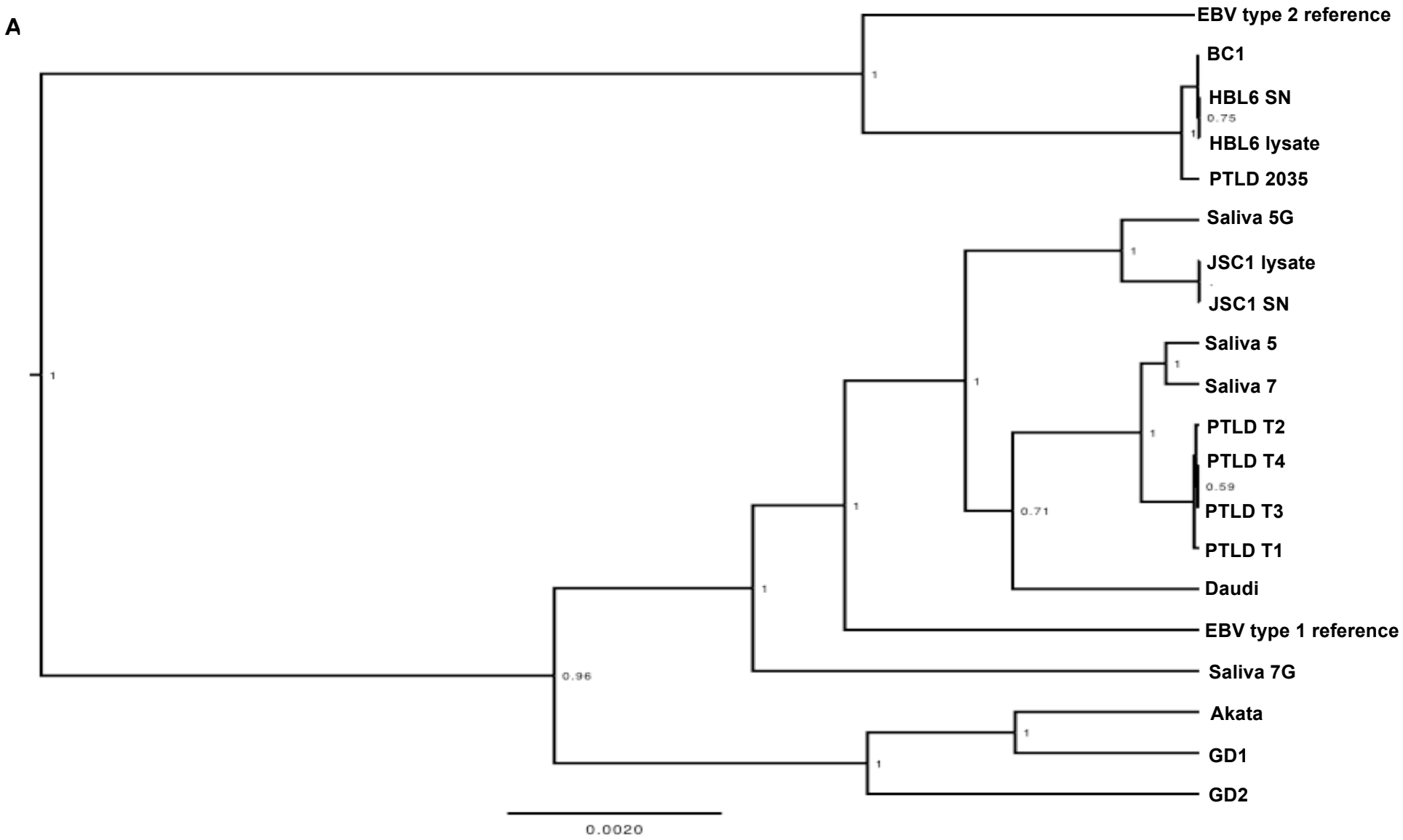
Repeat regions were removed and several methods were used to construct the phylogeny of the whole EBV genome (WG tree). Firstly, the Bayesian skyline coalescent model (Drummond & Rambaut, 2007; Drummond *et al.*, 2005) (referred to as Bayesian model) is used, which was regarded as the model with least restriction. The most suitable heterogeneity site model was first examined using JModelTest (Posada, 2008), and GTR substitution model with both gamma and invariant heterogeneity model were used for the analysis. Two WG trees were built using either relax longitudinal or exponential clock models, with

the sampling length ranging between 20 and 50 million. We also generated a distance based tree using neighbour joining phylogeny, which is a bottom-up clustering method generating an unrooted tree if no out-group is used. All methods generated the same tree structure. Figure 5.6.A shows the WG phylogenetic tree constructed by Bayesian method, using relax longitudinal model. Recombination between the samples was also examined using RDP3; however we could find no consistent evidence of recombination within these samples.

It has been reported that in HCMV variation occurs in higher rate at the CDS region in comparison to the non-CDS region (Renzette *et al.*, 2011). We have observed a higher number of SNPs at the CDs in comparison to the non-CDS region in EBV genomes from both cell lines and PTLD samples. Another phylogenetic tree was constructed using the same method, but with coding region only (repeats removed) sequences from the 20 whole genome sequences. The location of the coding region is determined *via* the annotation of the EBV type 1 reference genome in GenBank. The EBV type 2 reference genome contains only 82 genes in comparison to the type 1 reference genome in Genbank; however, the CDS sequence of the type 2 samples were generated using annotation of EBV type 1, since the annotation of the repeated regions and CDS of type 2 are not as extensively studied.

Figure 5.6.B shows the coding regions only of the EBV genomes phylogenetic tree (refer to as the CDS tree), constructed using the same methods as the WG tree. These two trees show the same grouping of the EBV genomes. Two major clades separate EBV type 1 and EBV type 2 samples. In the type 2 clade, unsurprisingly, BC1 EBV shows close association with HBL6 EBV sequences. BC1 and HBL6 are two independently established PEL cell lines from the same PEL tumor (Cesarman *et al.*, 1995b; Gaidano *et al.*, 1996). This further illustrates the robustness of our sequencing and analysis pipeline. In the type 1 clade, there is no evidence of EBV sequences grouping by the disease. EBV sequences from saliva samples are in the same clade as EBV from PEL, BL and PTLD. GD1 and GD2 sequences show higher diversity and are clustered away from rest of the type 1 sequences, but together with the Akata EBV sequence. The Akata cell line is derived from tumour cells of a Japanese BL patient (Takada, 1984; Takada *et al.*, 1991). Notably all three GD1, GD2 and

Akata EBV genomes were derived from samples originated from Asia, with GD1 and GD2 derived from Chinese patients with NPC, suggesting the possibility of geographical association in this clade.


A
EBV type 2 reference
BC1
HBL6 SN
0.75
HBL6 lysate
PTLD 2035
Saliva 5G
JSC1 lysate
JSC1 SN
Saliva 5
Saliva 7
PTLD T2
PTLD T4
0.59
PTLD T3
PTLD T1
Daudi
EBV type 1 reference
Saliva 7G
Akata
GD1
GD2
0.71
0.96
0.0020

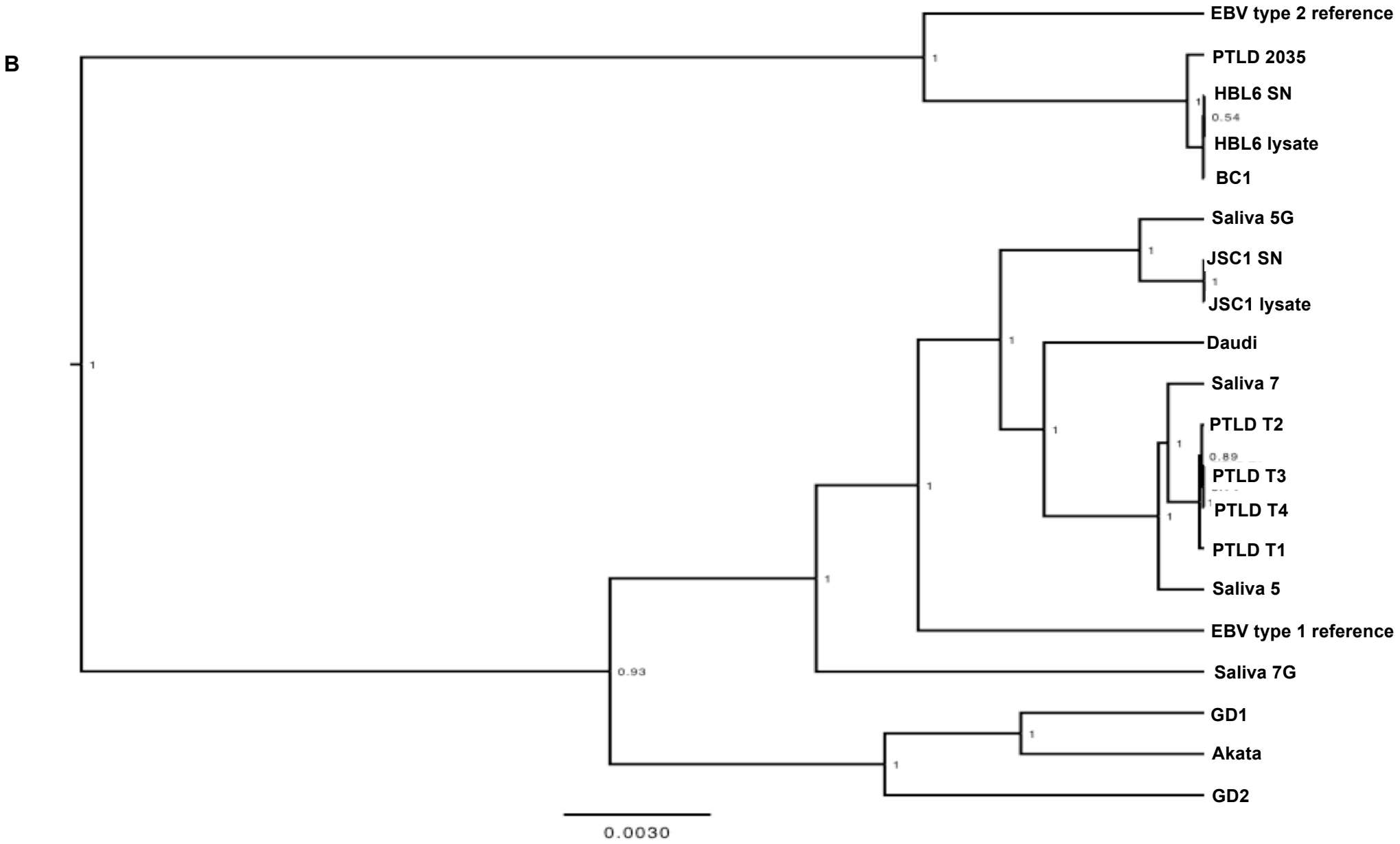

B
EBV type 2 reference
PTLD 2035
HBL6 SN
0.54
HBL6 lysate
BC1
Saliva 5G
JSC1 SN
JSC1 lysate
Daudi
Saliva 7
PTLD T2
0.89
PTLD T3
PTLD T4
PTLD T1
Saliva 5
EBV type 1 reference
Saliva 7G
GD1
Akata
GD2
0.93
0.0030

Figure 5.6. Phylogenetic tree of EBV A) whole genome sequences with no repeat regions, B) EBV CDS post removing the repeats. The tree is constructed using Bayesian skyline coalescent model, with relax longitudinal clock rooting of the tree. The EBV type 2 sequences: HBL6, BC1, PTLD 2035 and the type 2 reference genomes cluster together, away from other EBV type 1 sequences samples. For the EBV type 1 samples, there is no clear pattern of EBV sequences according to the type of diseases caused, especially with Akata within the same clade as GD1 and GD2. However, the numbers of samples is too small to draw any firm conclusion. PTLD samples T1 to T4 shows progression in time, with T1 differing more than the other three samples. (SN = supernatant). The divergence scale (indicating numbers of substitution per site) is shown at the bottom of the tree.

5.2.9. Validation of the majority consensus sequence *via* PCR and Sanger capillary sequencing

To validate the majority consensus sequence (MCS) assembled by QUASR v7, Sanger capillary sequencing (SCS) was carried out to sequence several regions across the EBV genome. The result of SCS was then compared to the MCS. Regions across two sections of BRLF1 gene, and one section each from BMRF1, BZLF1 and BDLF3.5 genes were cloned and table 5.6 shows the number of mis-matches between these two groups of sequences in these regions. SCS was also used to validate the EBV MCS of the PTLD samples across the EBNA2 gene. However, only PTLD T1 and T3 samples were sequenced due to the small volume of the original samples.

Table 5.6. Numbers of mis-matched bases between the majority consensus sequence and the PCR sequence.

	BRLF1		BMRF1	BZLF1	BDLF3.5	EBNA2
	section 1	section 2				
Daudi	0	0	N/A	0	0	N/A
Akata	0	1/843	0	29/611	0	N/A
HBL6	0	0	0	0	0	N/A
JSC1	0	N/A	N/A	0	0	N/A
PTLD T1	N/A	N/A	N/A	N/A	N/A	0
PTLD T3	N/A	N/A	N/A	N/A	N/A	0

For the cell line samples, a single clone of the PCR product was Sanger capillary sequenced (only one colony was picked for the cell line samples). There were no mis-matches between the DNA sequences generated by either SCS or MGC in EBV genomes of Daudi, HBL6 and JSC1 cell lines. However, there was a mis-match at the BRLF1 region that had been sequence in the EBV genome from the Akata cell line. The PCR product contains 843 bases of DNA sequence cloned from the supernatant of Akata cell line. The PCR sequence differs from the MCS by 1 substitution. Further inspection of the VarScan analysis of the MCS at the particular base indicates no minority population and adequate coverage.

Another difference between the SCS and MCS sequences also occurred in the Akata EBV genome, over the region of the BZLF1 gene (table 5.5). A total of 611 bases of BZLF-1 were sequenced, and the cloned PCR sequence showed a 29bp insertion (figure 5.7). The BZLF1 gene consists of three exons and one intron, which contains a section of 29 bps repeats. Previous studies have shown that there are three copies of the 29bp repeats in the EBV of Akata cell lines, with only two copies in the B95.8 cell line. B95.8 forms the majority of the EBV reference genome (Packham *et al.*, 1993). The primers designed to clone the BZLF1 region cover the intron including the repeats. This highlights a known problem in reference-based NGS assembly. As the majority consensus sequence of Akata EBV was produced by assembly against the EBV type 1 (B95.8) reference genome, the mapping algorithm assigns the number of repeats of the reference genome, (two copies of the 29 bps repeats), rather than the correct number of repeats as determined by SCS, namely three repeats.

Regions of the EBNA2 gene of PTLD T1 and T3 were also been cloned and sequenced by SCS. There were no mis-matches between the cloned sequence and the MCS of PTLD T1 and T3. Together this shows generating EBV consensus sequence from the Next Generation Sequencing product by reference based mapping has been effective and representative, but that care should be taken interpreting repeats in the genome.

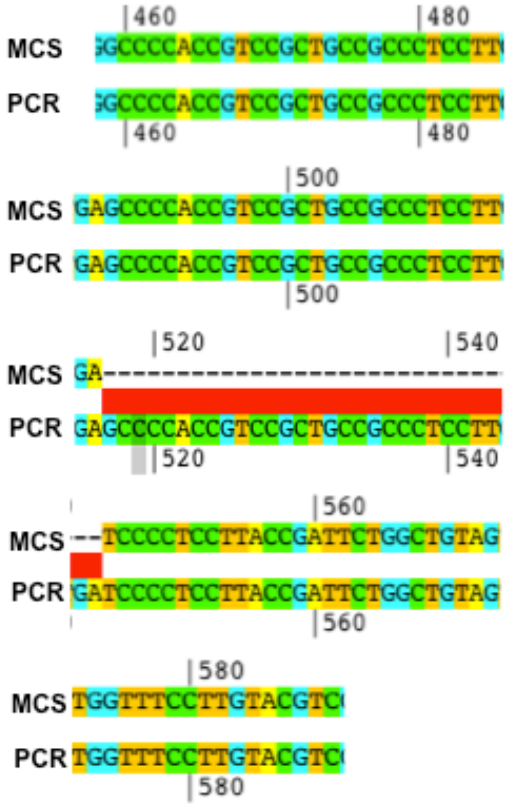


Figure 5.7. Comparison between regions of majority consensus sequence (MCS) and the cloned PCR product sequence of the EBV from Akata cell line. MCS contains two copies of the 29 bp repeats which resemble the EBV type 1 reference genome, with the PCR product showing three copies of the repeats. The difference was due to the reference based mapping method. MCS was constructed by mapping the NGS read pairs back to the B95.8 reference genome, therefore the number of repeats in the MCS is restricted to be the same as the reference genome, and does not represent the real number of repeats in the genome.

5.3. Conclusions

EBV whole genome analysis has not been possible previously due to a lack of whole genome data and a robust and cheap method to generate whole genomes. As discussed in Chapter 4, a pipeline has been developed which allows sequencing of EBV whole genomes from various types of samples using Next Generation Sequencing. So far, we have successfully sequenced and assembled EBV genomes from both cell lines and blood samples, and of both EBV type 1 and type 2.

Here, we investigate the difference between the EBV genomes. First EBV genomes from latency and the lytic cycle have been sequenced from both JSC1 and HBL6 cell lines, both are PEL cell lines dually infected with KSHV and EBV. The EBV genomes, with the repeated region of the genome removed, show 100% consistency between the latent and lytic cycle EBV sequences. This shows our method is robust and suggests that the EBV episomes in both cell lines are clonal.

The numbers of SNPs in EBV genomes across different types of samples were examined. SNPs occur every 170-360bp on average, with between 26-30% of the SNPs being non-synonymous. Interestingly, all the samples show similar locations of the NS-SNPs. Further analysis of the non-synonymous SNPs with the PTLD EBV samples showed a more uneven distribution across the genes. The majority of the NS-SNPs for the PTLD EBV sequence occured at the tegument and glycoprotein encoded genes, with EBNA1, 2 and LMP1 genes also polymorphic. However, EBNA1 contained the highest number of NS-SNPs in the genome of both the PTLD and three EBV sequences from cell lines, indicating major diversity in EBNA1 genes. EBNA1 is a DNA binding protein responsible for maintaining episomal DNA during cell division, previously shown to be highly polymorphic. During latency I, EBNA1 was the only viral protein expressed. EBNA1 contains a central repeat region composed of glycine-alanine repeats (GAr), which inhibits cytotoxic T lymphocyte (CTL) immune responses. The EBNA1 GAr domain inhibits proteasomal degradation of the EBNA1 protein (Levitskaya *et al.*, 1995) although the mechanism is unclear. EBNA1 also inhibits CTL peptide antigen processing, with GAr domain delaying the translation of EBNA1 and increasing the stability and reducing unfolded

protein turnover (Yin *et al.*, 2003). Further investigation of the mutations in the EBNA1 gene in the one PTLD is required to understand their cause. It is possible that immune deficiency in PTLD allows CTL escape variants to replicate that would normally be actively controlled. The EBV whole genome sequenced from PTLD patients were from blood samples, it would be interesting to see the EBV whole genome sequences in the PTLD tumour tissue sample. Other highly polymorphic genes from EBV genomes include the BFLF2, BNRF1, EBNA2, BPLF1 and BOLF1 genes. The functions of the majority of these genes are unknown and further investigation is required to understand the relationship between the genes and polymorphisms. BNRF1 encodes a tegument protein, and has been shown to activate early viral gene transcription (Lopez *et al.*, 2005), and also to allow efficient transfer of the viral DNA from endosomal compartments to the nucleus in B-cells (Tsai *et al.*, 2011). However, it is unclear what the effect of the variations of BNLF1 are on its function. BPLF1 shares structure homology with HSV1 VP16 protein which is a structural tegument component and is important for HSV replication and a trans-activator of viral immediate-early genes (Schmaus *et al.*, 2004). HSV VP16 protein forms a transcriptional regulatory complex with two cellular proteins: Oct-1, a POU-domain transcription factor, and HCF-1, a cell-proliferation factor upon infection, to activate the transcription of the first set of expressed viral genes (Wysocka & Herr, 2003). Recent studies have suggested that Oct-1 interacts with EBV BRLF1 protein and acts as a positive regulator of EBV lytic gene expression (Robinson *et al.*, 2011). Another POU-domain transcription factor Oct-2 has also been shown to maintain EBV latency by inhibiting EBV BZLF1 activation, possibility *via* direct binding to the BZLF1 protein (Robinson *et al.*, 2012). It is possible that BPLF1 gene, through sharing functional similarity to HSV VP16, also plays important role in EBV reactivation and infection.

We also investigated the changes of EBV sequences in a PTLD patient over a two month period. There has been no previous study showing the changes of EBV genome sequences through time. The EBV whole genomes assembled are the majority consensus sequences, where SNPs are called when there is a 100% change of base relative to the reference. We also examined the alteration of relative nucleotide levels between different EBV samples. EBNA1 and other highly polymorphic genes remain constant and do not change much over the

two month period of sampling. The BOLF1 gene that encodes a tegument protein is the most variable gene through the time course. The function of the BOLF1 gene is unclear, although it is suspected that the protein is responsible for virus assembly and budding. The reason for a high variation rate of one gene is not known.

After closer examination of the PTLD majority sequences over time, we have also observed alteration in the percentage of nucleotides present as NS-SNPs, indicating changes in minority species over time. The majority of the minority species of NS-SNPs occur in the BOLF1 gene. Since EBV is polyclonal in most of the PTLD patients, the alteration of minority species provides evidence of virus evolution throughout disease progression. The variation in minority species could also indicate the polyclonal virus population in the PTLD blood samples. However, with EBNA1 being the most polymorphic gene in the EBV genome, and remaining constant in this study, this suggests the variation of minority species in PTLD blood samples is due to evolutional pressure other than changes of virus population. However, further analysis of minority species in PTLD patients is required and since the samples were processed with WGA and also SureSelect process prior to NGS, it is unclear if these combined processes introduced a bias.

Phylogenetic trees have been constructed based on both whole genome and the CDS sequences. In both cases the phylogeny is identical and shows distinct separation between EBV type 1 and type 2 sequences. No clear separation relating to EBV disease can be seen from the tree, indicating no evolutionary differentiation of EBV genomes leading to different malignancies. However, the numbers of samples are too small for any firm conclusions. Liu *et al.,* (2011) had shown that for the EBNA1, BZLF1 and LMP1 genes, GD1 and GD2 consistently clustered together and share the same ancestor as the AG876 EBV sequence (EBV type 2 reference genome). Here we show that using every CDS of the EBV genomes, GD1 and GD2 indeed cluster together, but are more closely related to EBV type 1 genomes than the EBV type 2 reference genome. The close grouping between GD1, GD2 and Akata EBV genomes in the phylogenetic tree, although not indicating the association of EBV genomes according to the type of disease, does suggest grouping by geographical origin. The Akata cell line was derived from a Japanese patient with BL, with both GD1

and GD2 sequenced from Chinese patients with NPC, these three EBV genomes therefore represent EBV originating from Asia. Further studies with more EBV sequences from different diseases and different geographical regions is required to understand more about the relationship between the virus and the diseases it is associated with.

Chapter 6

General Discussion

EBV is one of the most widely distributed virus infections with 95% of the world population infected. EBV is linked to several different types of diseases, but little is known about virus reactivation *in vivo* and the diversity of whole EBV genome sequences.

The different latent and lytic cycles are one of the most important features of the herpesvirus life cycle. Latency allows EBV to remain hidden from the immune system and maintain a persistent infection, whilst lytic replication allows viruses to amplify and infect more target cells. The balance between these two phases is maintained and facilitated by many cellular factors to ensure the virus only reactivates under the correct conditions.

B-cell terminal differentiation into plasma cells has been associated with the reactivation of gamma herpesviruses. XBP-1, a transcriptional factor that is important for the UPR response during plasma cell differentiation has been shown to transactivate the ORF50 promoter and initiate the lytic cycle of KSHV (Dalton-Griffin *et al.*, 2009; Wilson *et al.*, 2007). Reactivation of other herpesviuses have also been linked to plasma cell differentiation, such as the murine gamma-herpesvirus 68 (Collins *et al.*, 2009) and EBV (Laichalk & Thorley-Lawson, 2005). However, studies on EBV lytic reactivation by XBP-1s have shown different results, with XBP-1s alone apparently able to transactivate the EBV BZLF1 gene and induce the lytic cycle (Sun & Thorley-Lawson, 2007); in contrast to another study indicating that XBP-1s requires PKD to be able to induce EBV BZLF1 protein expression and initiate the EBV lytic cycle (Bhende *et al.*, 2007).

Here we have shown that XBP-1s does not transactivate EBV IE gene promoters and does not induce BZLF1 or BRLF1 mRNA and protein expression, with or without PKD expression in tumour cell lines. The chemical inducing agents TPA and NaB are able to induce EBV lytic reactivation; however, they do not do so by inducing XBP-1 splicing in B-cells. Multiple myeloma cell lines constantly express XBP-1s, and in a recent study

superinfection of multiple myeloma cell lines by EBV showed no signs of constant lytic replication of EBV (Anastasiadou *et al.*, 2009). These results all indicate that XBP-1s expression does not induce EBV lytic reactivation in B-cells, contradicting the previous literature. The differences could be due to the cell type used in the experiments, and also the different strains of XBP-1s. Murine XBP-1s were used in previous studies (Bhende *et al.*, 2007; Sun & Thorley-Lawson, 2007), whereas human XBP-1s were used in this PhD study. EMSA were performed previously to identify the putative XBP-1s binding site on the Zp (Sun & Thorley-Lawson, 2007). Two putative binding sites Zp1 and Zp2 were identified using computational program, with Zp1 showing interaction with XBP-1s in the EMSA. However, Zp1 does not contain a similar sequence to the XBP-1 binding site on KSHV RTA promoter, whereas Zp2 shows homology. A CpG motif was identified in the Zp2 region. Zp does not interact with BZLF1, the protein that is able to self-activate, unless in the methylation status. The methylated state of Zp therefore is a factor in reactivation and may be the reason for the XBP-1s binding site in Zp being non-functional.

EBV has been associated with many different types of diseases. However, the whole genome sequences of EBV in both disease and health have not been studied. As with all other viruses it is important to study the diversity of the EBV genome across different diseases and geographical areas. Whole EBV genome sequencing has been difficult to carry out due to the nature of host DNA association of gamma herpesvirus. We have used the Agilent SureSelect enrichment system to enrich the viral DNA and remove the non-viral DNA from the samples. A pipeline to sequence EBV using NGS and computational assembly of the majority consensus sequence has been established, allowing faster and cheaper methods of generating EBV whole genome sequences. In this thesis, we have generated the majority consensus sequences of EBV from BLs and PEL cell lines and from clinical blood samples from PTLD patients.

Here we have shown that the EBV genome contains more variation than previously recognised. DNA viruses have not been linked to a high rate of variation; however, we have shown that the SNPs in EBV occur every 170-360bp within the genome, excluding the repeat regions. With the high variation between the genomes, it is debatable that the current reference sequence of EBV is the best representative of the virus genome. The non-synonymous

SNPs of EBV occur mostly in the glycoprotein and tegument proteins, although also in EBNA1, 2 and LMP1 genes. The EBNA1 gene is known to contain high amount of variation, and here we also shown it contains the highest amount of non-synonymous SNPs in different EBV majority consensus genomes. The effect of the SNPs in the EBNA1 gene is unclear. Other highly polymorphic genes that have not been identified previously include BOLF1, BPLF1 and BNRF1. BPLF1 shares structural similarity with HSV VP16 protein, which is important for HSV replication and transactivates the HSV IE gene. BNRF1 has recently been shown to activate early viral gene transcription and allows viral DNA to be transferred from the endosomal compartment to the nucleus. BOLF1 has been shown to be highly polymorphic in both EBV from cell lines and also PTLD samples, and also has a high mutation rate over the two months of the longitudinal EBV study. However, little is known about BOLF1 and the role it plays in the EBV life cycle and infectivity is unclear.

The phylogenetic study shows no clear association between the EBV sequences and their associated disease, despite a clear separation between type 1 and type 2 EBV genomes. The EBV genomes originating in Asia form a sub-group within the EBV type 1 clade, indicating a possibility of geographical clustering. However, the sampling number and geographical origin is too small to draw any firm conclusion from the study.

EBV whole genome sequences analysis has shown interesting variations throughout the genome. The variations highlighted in this study serve as a beginning of understanding EBV evolution and the difference between EBV populations in different geographical regions and diseases.

Chapter 7

Future Experiments

The thesis can be divided into two sections, in the first section we have examined the reactivation of EBV in the tumour B-cell background and in the second section we have developed a pipeline of sequencing EBV whole genome sequence using Next Generation Sequencing and investigated variation of between EBV genomes.

Further investigation of EBV reactivation in mature B-cells is required to understand the process fully. A recent study has shown ER stress induces EBV reactivation (Taylor *et al.*, 2011). During terminal differentiation, the germinal centre B-cells expand the secretary apparatus in preparation for producing large amount of antibodies, which induce ER stress. Other cellular factors involved in the ER stress response should be investigated in relation to EBV reactivation. TPA can reactivate both KSHV and EBV in B-cells into the lytic cycle *via* an XBP-1 independent pathway since TPA does not induce XBP-1 splicing. Recent studies have shown TPA dependent KSHV reactivation uses MEK/ERK, JNK and p38 signally pathways (Ford *et al.*, 2006; Xie *et al.*, 2008). As TPA also induces EBV reactivation, it would be interesting to carry out further investigations on these pathways and their effect on EBV reactivation.

EBV can be induced into lytic reactivation by histone deacetylate inhibitors such as sodium butyrate and valporic acid. MEF2D is known to recruit type II histone de-acetylase to the Zp and prevent EBV reactivation. Oct-2, a POU family protein has also recently been shown to be important for maintaining EBV latency in B-cells. It would be interesting to investigate if EBV sould enter lytic replication if these two proteins were disrupted, which could be achieved by knocking down these two proteins by using RNA interference (RNAi).

The second part of the thesis concentrates on the EBV whole genome sequencing. A pipeline of generating EBV whole genome majority consensus sequence has been established in this thesis. This pipeline can be used to generate large numbers of EBV whole genome sequences. However, there are several aspects of this pipeline that can be improved. We have shown that using the SureSelect system allows us to capture the target EBV DNA

fragments successfully. With more understanding of the EBV genome, the probes used in the capturing process could be re-designed to ensure higher efficiency. The WGA process has been shown to bias towards amplifying host DNA and therefore this step should be removed from the pipeline. Samples with low DNA content should either be processed by hand or using a carrier DNA to increase the total DNA content of the sample, which would allow the samples to be processed in the SureSelect procedure.

Currently the EBV whole genome is assembled against the reference sequences, resulting in a majority consensus sequences. Further investigation of compiling these sequences using *de novo* assembly is required, and the comparison between these two methods is necessary to ensure the accuracy of the EBV sequences.

With the highly variable EBNA1 genes from the PTLD samples, further analysis of the protein structure is required to understand, firstly if the variations alter the ability of EBNA1 to bind DNA, and therefore affect its ability of associating EBV episome with human DNA; and secondly if the variations affect the ability of EBNA1 to evade immune surveillance. Further investigation of the two genes that vary the most during the PTLD time series study are also needed, especially the BPLF1 and BOLF1 genes. Since BPLF1 shares structural homology with the HSV VP16 protein, it would be interesting to investigate the high rate of variation over time. Since limited studies have been carried out, further investigation of BPLF1 and the role it plays in EBV reactivation are required, especially its association with Oct-1. Limited studies have also been carried out on the BOLF1 gene, which encodes for a tegument protein. Since the BOLF1 gene shares structural homology with HLA-DQw8 beta chain, it would be interesting to further investigate the role of BOLF1 in host immune response.

To understand the link between EBV whole genome variation and the diseases associated with EBV, more EBV genome sequences from different diseases and healthy individuals in different geographical regions are needed.

References

Abbot, S. D., Rowe, M., Cadwallader, K., Ricksten, A., Gordon, J., Wang, F., Rymo, L. & Rickinson, A. B. (1990). Epstein-Barr virus nuclear antigen 2 induces expression of the virus-encoded latent membrane protein. *J Virol* **64**, 2126-2134.

Acosta-Alvear, D., Zhou, Y., Blais, A., Tsikitis, M., Lents, N. H., Arias, C., Lennon, C. J., Kluger, Y. & Dynlacht, B. D. (2007). XBP1 controls diverse cell type- and condition-specific transcriptional regulatory networks. *Mol Cell* **27**, 53-66.

Adams, A. (1987). Replication of latent Epstein-Barr virus genomes in Raji cells. *J Virol* **61**, 1743-1746.

Adams, B., Dorfler, P., Aguzzi, A., Kozmik, Z., Urbanek, P., Maurer-Fogy, I. & Busslinger, M. (1992). Pax-5 encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis. *Genes Dev* **6**, 1589-1607.

Adamson, A. L., Darr, D., Holley-Guthrie, E., Johnson, R. A., Mauser, A., Swenson, J. & Kenney, S. (2000). Epstein-Barr virus immediate-early proteins BZLF1 and BRLF1 activate the ATF2 transcription factor by increasing the levels of phosphorylated p38 and c-Jun N-terminal kinases. *J Virol* **74**, 1224-1233.

Adessi, C., Matton, G., Ayala, G., Turcatti, G., Mermod, J. J., Mayer, P. & Kawashima, E. (2000). Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. *Nucleic Acids Res* **28**, E87.

Adler, B., Schaadt, E., Kempkes, B., Zimber-Strobl, U., Baier, B. & Bornkamm, G. W. (2002). Control of Epstein-Barr virus reactivation by activated CD40 and viral latent membrane protein 1. *Proc Natl Acad Sci U S A* **99**, 437-442.

Ahearn, J. M., Hayward, S. D., Hickey, J. C. & Fearon, D. T. (1988). Epstein-Barr virus (EBV) infection of murine L cells expressing recombinant human EBV/C3d receptor. *Proc Natl Acad Sci U S A* **85**, 9307-9311.

Alfieri, C., Birkenbach, M. & Kieff, E. (1991). Early events in Epstein-Barr virus infection of human B lymphocytes. *Virology* **181**, 595-608.

Ali, A. K., Saito, S., Shibata, S., Takada, K. & Kanda, T. (2009). Distinctive effects of the Epstein-Barr virus family of repeats on viral latent gene promoter activity and B-lymphocyte transformation. *J Virol* **83**, 9163-9174.

Allan, G. J., Inman, G. J., Parker, B. D., Rowe, D. T. & Farrell, P. J. (1992). Cell growth effects of Epstein-Barr virus leader protein. *J Gen Virol* **73 (Pt 6)**, 1547-1551.

Allday, M. J., Crawford, D. H. & Griffin, B. E. (1989). Epstein-Barr virus latent gene expression during the initiation of B cell immortalization. *J Gen Virol* **70 (Pt 7)**, 1755-1764.

Allday, M. J. & Farrell, P. J. (1994). Epstein-Barr virus nuclear antigen EBNA3C/6 expression maintains the level of latent membrane protein 1 in G1-arrested cells. *J Virol* **68**, 3491-3498.

Allen, C. D., Ansel, K. M., Low, C., Lesley, R., Tamamura, H., Fujii, N. & Cyster, J. G. (2004). Germinal center dark and light zone organization is mediated by CXCR4 and CXCR5. *Nat Immunol* **5**, 943-952.

Allman, D., Li, J. & Hardy, R. R. (1999). Commitment to the B lymphoid lineage occurs before DH-JH recombination. *J Exp Med* **189**, 735-740.

Allman, D., Lindsley, R. C., DeMuth, W., Rudd, K., Shinton, S. A. & Hardy, R. R. (2001). Resolution of three nonproliferative immature splenic B cell subsets reveals multiple selection points during peripheral B cell maturation. *J Immunol* **167**, 6834-6840.

Allman, D. & Pillai, S. (2008). Peripheral B cell subsets. *Curr Opin Immunol* **20**, 149-157.

Allman, D. M., Ferguson, S. E., Lentz, V. M. & Cancro, M. P. (1993). Peripheral B cell maturation. II. Heat-stable antigen(hi) splenic B cells are an immature developmental intermediate in the production of long-lived marrow-derived B cells. *J Immunol* **151**, 4431-4444.

Almqvist, J., Zou, J., Linderson, Y., Borestrom, C., Altiok, E., Zetterberg, H., Rymo, L., Pettersson, S. & Ernberg, I. (2005). Functional interaction of Oct transcription factors with the family of repeats in Epstein-Barr virus oriP. *J Gen Virol* **86**, 1261-1267.

Ambinder, R. F., Shah, W. A., Rawlins, D. R., Hayward, G. S. & Hayward, S. D. (1990). Definition of the sequence requirements for binding of the EBNA-1 protein to its palindromic target sites in Epstein-Barr virus DNA. *J Virol* **64**, 2369-2379.

Amon, W., Binne, U. K., Bryant, H., Jenkins, P. J., Karstegl, C. E. & Farrell, P. J. (2004). Lytic cycle gene regulation of Epstein-Barr virus. *J Virol* **78**, 13460-13469.

Amon, W. & Farrell, P. J. (2005). Reactivation of Epstein-Barr virus from latency. *Rev Med Virol* **15**, 149-156.

Anastasiadou, E., Vaeth, S., Cuomo, L., Boccellato, F., Vincenti, S., Cirone, M., Presutti, C., Junker, S., Winberg, G., Frati, L., Wade, P. A., Faggioni, A. & Trivedi, P. (2009). Epstein-Barr virus infection leads to partial phenotypic reversion of terminally differentiated malignant B cells. *Cancer Lett* **284**, 165-174.

Anderlini, P., Valbuena, J. R., Champlin, R. E. & Bueso-Ramos, C. E. (2004). Epstein-Barr virus-associated, CD20- polyclonal lymphoproliferative disorder after matched unrelated donor marrow transplantation. *Bone Marrow Transplant* **34**, 919-921.

Andersson-Anvret, M., Forsby, N., Klein, G., Henle, W. & Biorklund, A. (1979). Relationship between the Epstein-Barr virus genome and nasopharyngeal carcinoma in Caucasian patients. *Int J Cancer* **23**, 762-767.

Apolloni, A. & Sculley, T. B. (1994). Detection of A-type and B-type Epstein-Barr virus in throat washings and lymphocytes. *Virology* **202**, 978-981.

Arpin, C., Banchereau, J. & Liu, Y. J. (1997). Memory B cells are biased towards terminal differentiation: a strategy that may prevent repertoire freezing. *J Exp Med* **186**, 931-940.

Arrand, J. R., Young, L. S. & Tugwood, J. D. (1989). Two families of sequences in the small RNA-encoding region of Epstein-Barr virus (EBV) correlate with EBV types A and B. *J Virol* **63**, 983-986.

Arvanitakis, L., Mesri, E. A., Nador, R. G., Said, J. W., Asch, A. S., Knowles, D. M. & Cesarman, E. (1996). Establishment and characterization of a primary effusion (body cavity-based) lymphoma cell line (BC-3) harboring kaposi's sarcoma-associated herpesvirus (KSHV/HHV-8) in the absence of Epstein-Barr virus. *Blood* **88**, 2648-2654.

Ascherio, A. & Munch, M. (2000). Epstein-Barr virus and multiple sclerosis. *Epidemiology* **11**, 220-224.

Atanasiu, C., Deng, Z., Wiedmer, A., Norseen, J. & Lieberman, P. M. (2006). ORC binding to TRF2 stimulates OriP replication. *EMBO Rep* **7**, 716-721.

Babcock, G. J., Decker, L. L., Freeman, R. B. & Thorley-Lawson, D. A. (1999). Epstein-barr virus-infected resting memory B cells, not proliferating lymphoblasts, accumulate in the peripheral blood of immunosuppressed patients. *J Exp Med* **190**, 567-576.

Babcock, G. J., Decker, L. L., Volk, M. & Thorley-Lawson, D. A. (1998). EBV persistence in memory B cells in vivo. *Immunity* **9**, 395-404.

Baer, R., Bankier, A. T., Biggin, M. D., Deininger, P. L., Farrell, P. J., Gibson, T. J., Hatfull, G., Hudson, G. S., Satchwell, S. C., Seguin, C. & et al. (1984). DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature* **310**, 207-211.

Bashaw, J. M. & Yates, J. L. (2001). Replication from oriP of Epstein-Barr virus requires exact spacing of two bound dimers of EBNA1 which bend DNA. *J Virol* **75**, 10603-10611.

Basseri, S. & Austin, R. C. (2012). Endoplasmic reticulum stress and lipid metabolism: mechanisms and therapeutic potential. *Biochem Res Int* **2012**, 841362.

Bassing, C. H., Swat, W. & Alt, F. W. (2002). The mechanism and regulation of chromosomal V(D)J recombination. *Cell* **109 Suppl**, S45-55.

Batista, F. D., Iber, D. & Neuberger, M. S. (2001). B cells acquire antigen from target cells after synapse formation. *Nature* **411**, 489-494.

Baxendale, A. J., Dawson, C. W., Stewart, S. E., Mudaliar, V., Reynolds, G., Gordon, J., Murray, P. G., Young, L. S. & Eliopoulos, A. G. (2005). Constitutive activation of the CD40 pathway promotes cell transformation and neoplastic growth. *Oncogene* **24**, 7913-7923.

Bayliss, G. J. & Wolf, H. (1981). The regulated expression of Epstein-Barr virus. III. Proteins specified by EBV during the lytic cycle. *J Gen Virol* **56**, 105-118.

Beisel, C., Tanner, J., Matsuo, T., Thorley-Lawson, D., Kezdy, F. & Kieff, E. (1985). Two major outer envelope glycoproteins of Epstein-Barr virus are encoded by the same gene. *J Virol* **54**, 665-674.

Bell, M. J., Brennan, R., Miles, J. J., Moss, D. J., Burrows, J. M. & Burrows, S. R. (2008). Widespread sequence variation in Epstein-Barr virus nuclear antigen 1 influences the antiviral T cell response. *J Infect Dis* **197**, 1594-1597.

Ben-Bassat, H., Goldblum, N., Mitrani, S., Goldblum, T., Yoffey, J. M., Cohen, M. M., Bentwich, Z., Ramot, B., Klein, E. & Klein, G. (1977). Establishment in continuous culture of a new type of lymphocyte from a "Burkitt like" malignant lymphoma (line D.G.-75). *Int J Cancer* **19**, 27-33.

Benner, R., van Oudenaren, A. & de Ruiter, H. (1977). B memory cells in the thymus: part of the pool of potentially circulating memory cells. *J Immunol* **119**, 1846-1848.

Beral, V., Peterman, T. A., Berkelman, R. L. & Jaffe, H. W. (1990). Kaposi's sarcoma among persons with AIDS: a sexually transmitted infection? *Lancet* **335**, 123-128.

Berek, C., Berger, A. & Apel, M. (1991). Maturation of the immune response in germinal centers. *Cell* **67**, 1121-1129.

Bergbauer, M., Kalla, M., Schmeinck, A., Gobel, C., Rothbauer, U., Eck, S., Benet-Pages, A., Strom, T. M. & Hammerschmidt, W. (2010). CpG-methylation regulates a class of Epstein-Barr virus promoters. *PLoS Pathog* **6**, e1001114.

Bernales, S., McDonald, K. L. & Walter, P. (2006). Autophagy counterbalances endoplasmic reticulum expansion during the unfolded protein response. *PLoS Biol* **4**, e423.

Bernasconi, N. L., Onai, N. & Lanzavecchia, A. (2003). A role for Toll-like receptors in acquired immunity: up-regulation of TLR9 by BCR triggering in naive B cells and constitutive expression in memory B cells. *Blood* **101**, 4500-4504.

Bernasconi, N. L., Traggiai, E. & Lanzavecchia, A. (2002). Maintenance of serological memory by polyclonal activation of human memory B cells. *Science* **298**, 2199-2202.

Bertolotti, A., Zhang, Y., Hendershot, L. M., Harding, H. P. & Ron, D. (2000). Dynamic interaction of BiP and ER stress transducers in the unfolded-protein response. *Nat Cell Biol* **2**, 326-332.

Besnier, C., Takeuchi, Y. & Towers, G. (2002). Restriction of lentivirus in monkeys. *Proc Natl Acad Sci U S A* **99**, 11920-11925.

Bhat, N. M., Kantor, A. B., Bieber, M. M., Stall, A. M., Herzenberg, L. A. & Teng, N. N. (1992). The ontogeny and functional characteristics of human B-1 (CD5+ B) cells. *Int Immunol* **4**, 243-252.

Bhende, P. M., Dickerson, S. J., Sun, X., Feng, W. H. & Kenney, S. C. (2007). X-box-binding protein 1 activates lytic Epstein-Barr virus gene expression in combination with protein kinase D. *J Virol* **81**, 7363-7370.

Bhende, P. M., Seaman, W. T., Delecluse, H. J. & Kenney, S. C. (2004). The EBV lytic switch protein, Z, preferentially binds to and activates the methylated viral genome. *Nat Genet* **36**, 1099-1104.

Biggar, R. J., Whitby, D., Marshall, V., Linhares, A. C. & Black, F. (2000). Human herpesvirus 8 in Brazilian Amerindians: a hyperendemic population with a new subtype. *J Infect Dis* **181**, 1562-1568.

Binne, U. K., Amon, W. & Farrell, P. J. (2002). Promoter sequences required for reactivation of Epstein-Barr virus from latency. *J Virol* **76**, 10282-10289.

Bishop, G. A. & Hostager, B. S. (2001). B lymphocyte activation by contact-mediated interactions with T lymphocytes. *Curr Opin Immunol* **13**, 278-285.

Bleesing, J. J. & Fleisher, T. A. (2003). Human B cells express a CD45 isoform that is similar to murine B220 and is downregulated with acquisition of the memory B-cell marker CD27. *Cytometry B Clin Cytom* **51**, 1-8.

Blink, E. J., Light, A., Kallies, A., Nutt, S. L., Hodgkin, P. D. & Tarlinton, D. M. (2005). Early appearance of germinal center-derived memory B cells and plasma cells in blood after primary immunization. *J Exp Med* **201**, 545-554.

Blond-Elguindi, S., Cwirla, S. E., Dower, W. J., Lipshutz, R. J., Sprang, S. R., Sambrook, J. F. & Gething, M. J. (1993). Affinity panning of a library of peptides displayed on bacteriophages reveals the binding specificity of BiP. *Cell* **75**, 717-728.

Bodescot, M. & Perricaudet, M. (1986). Epstein-Barr virus mRNAs produced by alternative splicing. *Nucleic Acids Res* **14**, 7103-7114.

Bodescot, M., Perricaudet, M. & Farrell, P. J. (1987). A promoter for the highly spliced EBNA family of RNAs of Epstein-Barr virus. *J Virol* **61**, 3424-3430.

Bommiasamy, H., Back, S. H., Fagone, P., Lee, K., Meshinchi, S., Vink, E., Sriburi, R., Frank, M., Jackowski, S., Kaufman, R. J. & Brewer, J. W. (2009). ATF6alpha induces XBP1-independent expansion of the endoplasmic reticulum. *J Cell Sci* **122**, 1626-1636.

Booth, K., Burkitt, D. P., Bassett, D. J., Cooke, R. A. & Biddulph, J. (1967). Burkitt lymphoma in Papua, New Guinea. *Br J Cancer* **21**, 657-664.

Borza, C. M. & Hutt-Fletcher, L. M. (2002). Alternate replication in B cells and epithelial cells switches tropism of Epstein-Barr virus. *Nat Med* **8**, 594-599.

Borza, C. M., Morgan, A. J., Turk, S. M. & Hutt-Fletcher, L. M. (2004). Use of gHgL for attachment of Epstein-Barr virus to epithelial cells compromises infection. *J Virol* **78**, 5007-5014.

Boshoff, C. & Weiss, R. (2002). AIDS-related malignancies. *Nat Rev Cancer* **2**, 373-382.

Boyle, G. J., Michaels, M. G., Webber, S. A., Knisely, A. S., Kurland, G., Cipriani, L. A., Griffith, B. P. & Fricker, F. J. (1997). Posttransplantation lymphoproliferative disorders in pediatric thoracic organ recipients. *J Pediatr* **131**, 309-313.

Brady, G., Macarthur, G. J. & Farrell, P. J. (2008). Epstein-Barr virus and Burkitt lymphoma. *Postgrad Med J* **84**, 372-377.

Bray, F., Haugen, M., Moger, T. A., Tretli, S., Aalen, O. O. & Grotmol, T. (2008). Age-incidence curves of nasopharyngeal carcinoma worldwide: bimodality in low-risk populations and aetiologic implications. *Cancer Epidemiol Biomarkers Prev* **17**, 2356-2365.

Breuer, J., Grose, C., Norberg, P., Tipples, G. & Schmid, D. S. (2010). A proposal for a common nomenclature for viral clades that form the species varicella-zoster virus: summary of VZV Nomenclature Meeting 2008, Barts and the London School of Medicine and Dentistry, 24-25 July 2008. *J Gen Virol* **91**, 821-828.

Breuer, J., Quinlivan, M., Al Bassam, M., Macdonald, S., Nichols, R. A., Steinberg, S. P., Russa, P. L. & Gershon, A. A. (2007). DNA sequence variability in Oka vaccine isolates. *J Infect Dis* **196**, 801-802; author reply 802-803.

Brodie, R., Smith, A. J., Roper, R. L., Tcherepanov, V. & Upton, C. (2004). Base-By-Base: single nucleotide-level analysis of whole viral genome alignments. *BMC Bioinformatics* **5**, 96.

Brooks, J. M., Croom-Carter, D. S., Leese, A. M., Tierney, R. J., Habeshaw, G. & Rickinson, A. B. (2000). Cytotoxic T-lymphocyte responses to a polymorphic Epstein-Barr virus epitope identify healthy carriers with coresident viral strains. *J Virol* **74**, 1801-1809.

Brooks, L., Yao, Q. Y., Rickinson, A. B. & Young, L. S. (1992). Epstein-Barr virus latent gene transcription in nasopharyngeal carcinoma cells: coexpression of EBNA1, LMP1, and LMP2 transcripts. *J Virol* **66**, 2689-2697.

Bryant, H. & Farrell, P. J. (2002). Signal Transduction and Transcription Factor Modification during Reactivation of Epstein-Barr Virus from Latency. *J Virol* **76**, 10290-10298.

Bu, W., Carroll, K. D., Palmeri, D. & Lukac, D. M. (2007). Kaposi's sarcoma-associated herpesvirus/human herpesvirus 8 ORF50/Rta lytic switch protein functions as a tetramer. *J Virol* **81**, 5788-5806.

Buck, M., Cross, S., Krauer, K., Kienzle, N. & Sculley, T. B. (1999). A-type and B-type Epstein-Barr virus differ in their ability to spontaneously enter the lytic cycle. *J Gen Virol* **80 (Pt 2)**, 441-445.

Burkhardt, A. L., Bolen, J. B., Kieff, E. & Longnecker, R. (1992). An Epstein-Barr virus transformation-associated membrane protein interacts with src family tyrosine kinases. *J Virol* **66**, 5161-5167.

Burkitt, D. (1958). A sarcoma involving the jaws in African children. *Br J Surg* **46**, 218-223.

Burkitt, D. (1962). A children's cancer dependent on climatic factors. *Nature* **194**, 232-234.

Burysek, L. & Pitha, P. M. (2001). Latently expressed human herpesvirus 8-encoded interferon regulatory factor 2 inhibits double-stranded RNA-activated protein kinase. *J Virol* **75**, 2345-2352.

Busch, L. K. & Bishop, G. A. (1999). The EBV transforming protein, latent membrane protein 1, mimics and cooperates with CD40 signaling in B lymphocytes. *J Immunol* **162**, 2555-2561.

Busson, P., Keryer, C., Ooka, T. & Corbex, M. (2004). EBV-associated nasopharyngeal carcinomas: from epidemiology to virus-targeting strategies. *Trends Microbiol* **12**, 356-360.

Cai, X., Lu, S., Zhang, Z., Gonzalez, C. M., Damania, B. & Cullen, B. R. (2005). Kaposi's sarcoma-associated herpesvirus expresses an array of viral microRNAs in latently infected cells. *Proc Natl Acad Sci U S A* **102**, 5570-5575.

Caldwell, R. G., Wilson, J. B., Anderson, S. J. & Longnecker, R. (1998). Epstein-Barr virus LMP2A drives B cell development and survival in the absence of normal B cell receptor signals. *Immunity* **9**, 405-411.

Calfon, M., Zeng, H., Urano, F., Till, J. H., Hubbard, S. R., Harding, H. P., Clark, S. G. & Ron, D. (2002). IRE1 couples endoplasmic reticulum load to secretory capacity by processing the XBP-1 mRNA. *Nature* **415**, 92-96.

Cannon, J. S., Ciufo, D., Hawkins, A. L., Griffin, C. A., Borowitz, M. J., Hayward, G. S. & Ambinder, R. F. (2000). A new primary effusion lymphoma-derived cell line yields a highly infectious Kaposi's sarcoma herpesvirus-containing supernatant. *J Virol* **74**, 10187-10193.

Carbone, A., Cilia, A. M., Gloghini, A., Capello, D., Fassone, L., Perin, T., Rossi, D., Canzonieri, V., De Paoli, P., Vaccher, E., Tirelli, U., Volpe, R. & Gaidano, G. (2000). Characterization of a novel HHV-8-positive cell line reveals implications for the pathogenesis and cell cycle control of primary effusion lymphoma. *Leukemia* **14**, 1301-1309.

Carbone, A. & Gloghini, A. (2008). KSHV/HHV8-associated lymphomas. *Br J Haematol* **140**, 13-24.

Carbone, A., Gloghini, A., Vaccher, E., Zagonel, V., Pastore, C., Dalla Palma, P., Branz, F., Saglio, G., Volpe, R., Tirelli, U. & Gaidano, G. (1996). Kaposi's sarcoma-associated herpesvirus DNA sequences in AIDS-related and AIDS-unrelated lymphomatous effusions. *Br J Haematol* **94**, 533-543.

Carel, J. C., Myones, B. L., Frazier, B. & Holers, V. M. (1990). Structural requirements for C3d,g/Epstein-Barr virus receptor (CR2/CD21) ligand binding, internalization, and viral infection. *J Biol Chem* **265**, 12293-12299.

Cariappa, A., Tang, M., Parng, C., Nebelitskiy, E., Carroll, M., Georgopoulos, K. & Pillai, S. (2001). The follicular versus marginal zone B lymphocyte cell fate decision is regulated by Aiolos, Btk, and CD21. *Immunity* **14**, 603-615.

Caselli, E., Galvan, M., Cassai, E., Caruso, A., Sighinolfi, L. & Di Luca, D. (2005). Human herpesvirus 8 enhances human immunodeficiency virus replication in acutely infected cells and induces reactivation in latently infected cells. *Blood* **106**, 2790-2797.

Casola, S., Otipoby, K. L., Alimzhanov, M., Humme, S., Uyttersprot, N., Kutok, J. L., Carroll, M. C. & Rajewsky, K. (2004). B cell receptor signal strength determines B cell fate. *Nat Immunol* **5**, 317-327.

Cattoretti, G., Chang, C. C., Cechova, K., Zhang, J., Ye, B. H., Falini, B., Louie, D. C., Offit, K., Chaganti, R. S. & Dalla-Favera, R. (1995). BCL-6 protein is expressed in germinal-center B cells. *Blood* **86**, 45-53.

Cesarman, E., Chang, Y., Moore, P. S., Said, J. W. & Knowles, D. M. (1995a). Kaposi's sarcoma-associated herpesvirus-like DNA sequences in AIDS-related body-cavity-based lymphomas. *N Engl J Med* **332**, 1186-1191.

Cesarman, E., Moore, P. S., Rao, P. H., Inghirami, G., Knowles, D. M. & Chang, Y. (1995b). In vitro establishment and characterization of two acquired immunodeficiency syndrome-related lymphoma cell lines (BC-1 and BC-2) containing Kaposi's sarcoma-associated herpesvirus-like (KSHV) DNA sequences. *Blood* **86**, 2708-2714.

Chadburn, A., Hyjek, E., Mathew, S., Cesarman, E., Said, J. & Knowles, D. M. (2004). KSHV-positive solid lymphomas represent an extra-cavitary variant of primary effusion lymphoma. *Am J Surg Pathol* **28**, 1401-1416.

Chan, J. K., Tsang, W. Y., Ng, C. S., Wong, C. S. & Lo, E. S. (1995). A study of the association of Epstein-Barr virus with Burkitt's lymphoma occurring in a Chinese population. *Histopathology* **26**, 239-245.

Chan, K. C. & Lo, Y. M. (2002). Circulating EBV DNA as a tumor marker for nasopharyngeal carcinoma. *Semin Cancer Biol* **12**, 489-496.

Chang, C. M., Yu, K. J., Mbulaiteye, S. M., Hildesheim, A. & Bhatia, K. (2009). The extent of genetic diversity of Epstein-Barr virus and its geographic and disease patterns: a need for reappraisal. *Virus Res* **143**, 209-221.

Chang, E. T. & Adami, H. O. (2006). The enigmatic epidemiology of nasopharyngeal carcinoma. *Cancer Epidemiol Biomarkers Prev* **15**, 1765-1777.

Chang, L. K. & Liu, S. T. (2000). Activation of the BRLF1 promoter and lytic cycle of Epstein-Barr virus by histone acetylation. *Nucleic Acids Res* **28**, 3918-3925.

Chang, Y., Cesarman, E., Pessin, M. S., Lee, F., Culpepper, J., Knowles, D. M. & Moore, P. S. (1994). Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science* **266**, 1865-1869.

Chaudhuri, B., Xu, H., Todorov, I., Dutta, A. & Yates, J. L. (2001). Human DNA replication initiation factors, ORC and MCM, associate with oriP of Epstein-Barr virus. *Proc Natl Acad Sci U S A* **98**, 10085-10089.

Chen, C. Z., Li, L., Lodish, H. F. & Bartel, D. P. (2004). MicroRNAs modulate hematopoietic lineage differentiation. *Science* **303**, 83-86.

Chen, H. L., Lung, M. L., Chan, K. H., Griffin, B. E. & Ng, M. H. (1996). Tissue distribution of Epstein-Barr virus genotypes. *J Virol* **70**, 7301-7305.

Chen, J., Ueda, K., Sakakibara, S., Okuno, T., Parravicini, C., Corbellino, M. & Yamanishi, K. (2001). Activation of latent Kaposi's sarcoma-associated herpesvirus by demethylation of the promoter of the lytic transactivator. *Proc Natl Acad Sci U S A* **98**, 4119-4124.

Chen, M. L., Wu, R. C., Liu, S. T. & Chang, Y. S. (1995). Characterization of 5'-upstream sequence of the latent membrane protein 1 (LMP-1) gene of an Epstein-Barr virus identified in nasopharyngeal carcinoma tissues. *Virus Res* **37**, 75-84.

Chene, A., Donati, D., Guerreiro-Cacais, A. O., Levitsky, V., Chen, Q., Falk, K. I., Orem, J., Kironde, F., Wahlgren, M. & Bejarano, M. T. (2007). A molecular link between malaria and Epstein-Barr virus reactivation. *PLoS Pathog* **3**, e80.

Cherayil, B. J. & Pillai, S. (1991). The omega/lambda 5 surrogate immunoglobulin light chain is expressed on the surface of transitional B lymphocytes in murine bone marrow. *J Exp Med* **173**, 111-116.

Chia, M. C., Leung, A., Krushel, T., Alajez, N. M., Lo, K. W., Busson, P., Klamut, H. J., Bastianutto, C. & Liu, F. F. (2008). Nuclear factor-Y and Epstein Barr virus in nasopharyngeal cancer. *Clin Cancer Res* **14**, 984-994.

Chou, L. S., Liu, C. S., Boese, B., Zhang, X. & Mao, R. (2010). DNA sequence capture and enrichment by microarray followed by next-generation sequencing for targeted resequencing: neurofibromatosis type 1 gene as a model. *Clin Chem* **56**, 62-72.

Cleary, M. L., Meeker, T. C., Levy, S., Lee, E., Trela, M., Sklar, J. & Levy, R. (1986). Clustering of extensive somatic mutations in the variable region of an immunoglobulin heavy chain gene from a human B cell lymphoma. *Cell* **44**, 97-106.

Cochet, C., Martel-Renoir, D., Grunewald, V., Bosq, J., Cochet, G., Schwaab, G., Bernaudin, J. F. & Joab, I. (1993). Expression of the Epstein-Barr virus immediate early gene, BZLF1, in nasopharyngeal carcinoma tumor cells. *Virology* **197**, 358-365.

Cohen, J. I., Wang, F., Mannick, J. & Kieff, E. (1989). Epstein-Barr virus nuclear protein 2 is a key determinant of lymphocyte transformation. *Proc Natl Acad Sci U S A* **86**, 9558-9562.

Collins, C. M., Boss, J. M. & Speck, S. H. (2009). Identification of infected B-cell populations by using a recombinant murine gammaherpesvirus 68 expressing a fluorescent protein. *J Virol* **83**, 6484-6493.

Cook, P. M., Whitby, D., Calabro, M. L., Luppi, M., Kakoola, D. N., Hjalgrim, H., Ariyoshi, K., Ensoli, B., Davison, A. J. & Schulz, T. F. (1999). Variability and evolution of Kaposi's sarcoma-associated herpesvirus in Europe and Africa. International Collaborative Group. *AIDS* **13**, 1165-1176.

Cotter, M. A., 2nd & Robertson, E. S. (2002). Molecular biology of Kaposi's sarcoma-associated herpesvirus. *Front Biosci* **7**, d358-375.

Countryman, J. & Miller, G. (1985). Activation of expression of latent Epstein-Barr herpesvirus after gene transfer with a small cloned subfragment of heterogeneous viral DNA. *Proc Natl Acad Sci U S A* **82**, 4085-4089.

Countryman, J. K., Gradoville, L. & Miller, G. (2008). Histone hyperacetylation occurs on promoters of lytic cycle regulatory genes in Epstein-Barr virus-infected cell lines which are refractory to disruption of latency by histone deacetylase inhibitors. *J Virol* **82**, 4706-4719.

Cox, K. L., Lawrence-Miyasaki, L. S., Garcia-Kennedy, R., Lennette, E. T., Martinez, O. M., Krams, S. M., Berquist, W. E., So, S. K. & Esquivel, C. O. (1995). An increased incidence of Epstein-Barr virus infection and lymphoproliferative disorder in young children on FK506 after liver transplantation. *Transplantation* **59**, 524-529.

Credle, J. J., Finer-Moore, J. S., Papa, F. R., Stroud, R. M. & Walter, P. (2005). On the mechanism of sensing unfolded protein in the endoplasmic reticulum. *Proc Natl Acad Sci U S A* **102**, 18773-18784.

Dales, S. & Chardonnet, Y. (1973). Early events in the interaction of adenoviruses with HeLa cells. IV. Association with microtubules and the nuclear pore complex during vectorial movement of the inoculum. *Virology* **56**, 465-483.

Dalldorf, G., Linsell, C. A., Barnhart, F. E. & Martyn, R. (1964). An Epidemiologic Approach to the Lymphomas of African Children and Burkitt's Sacroma of the Jaws. *Perspect Biol Med* **7**, 435-449.

Dalton-Griffin, L., Wilson, S. J. & Kellam, P. (2009). X-box binding protein 1 contributes to induction of the Kaposi's sarcoma-associated herpesvirus lytic cycle under hypoxic conditions. *J Virol* **83**, 7202-7209.

Dambaugh, T., Beisel, C., Hummel, M., King, W., Fennewald, S., Cheung, A., Heller, M., Raab-Traub, N. & Kieff, E. (1980). Epstein-Barr virus (B95-8) DNA VII: molecular cloning and detailed mapping. *Proc Natl Acad Sci U S A* **77**, 2999-3003.

Dean, F. B., Nelson, J. R., Giesler, T. L. & Lasken, R. S. (2001). Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res* **11**, 1095-1099.

Decker, L. L., Klaman, L. D. & Thorley-Lawson, D. A. (1996). Detection of the latent form of Epstein-Barr virus DNA in the peripheral blood of healthy individuals. *J Virol* **70**, 3286-3289.

Dedicoat, M. & Newton, R. (2003). Review of the distribution of Kaposi's sarcoma-associated herpesvirus (KSHV) in Africa in relation to the incidence of Kaposi's sarcoma. *Br J Cancer* **88**, 1-3.

Deng, H., Young, A. & Sun, R. (2000). Auto-activation of the rta gene of human herpesvirus-8/Kaposi's sarcoma-associated herpesvirus. *J Gen Virol* **81**, 3043-3048.

Deng, Z., Lezina, L., Chen, C. J., Shtivelband, S., So, W. & Lieberman, P. M. (2002). Telomeric proteins regulate episomal maintenance of Epstein-Barr virus origin of plasmid replication. *Mol Cell* **9**, 493-503.

Dent, A. L., Shaffer, A. L., Yu, X., Allman, D. & Staudt, L. M. (1997). Control of inflammation, cytokine expression, and germinal center formation by BCL-6. *Science* **276**, 589-592.

Depledge, D. P., Palser, A. L., Watson, S. J., Lai, I. Y., Gray, E. R., Grant, P., Kanda, R. K., Leproust, E., Kellam, P. & Breuer, J. (2011). Specific capture and whole-genome sequencing of viruses from clinical samples. *PLoS One* **6**, e27805.

Dickerson, S. J., Xing, Y., Robinson, A. R., Seaman, W. T., Gruffat, H. & Kenney, S. C. (2009). Methylation-dependent binding of the epstein-barr virus BZLF1 protein to viral promoters. *PLoS Pathog* **5**, e1000356.

Diebold, J. J., E.S.; Raphael, M.; WarnKe, R.A. (2001). Burkitt Lymphoma. In ***Pathology and Genetics: Tumours of Haematopoietic and Lymphoid Tissues***, pp. 181-184. Edited by N. L. H. Elaine S. Jaffe, Harald Stein, James W Vardiman: IARC Press, International Agency for REsearch on Cancer (IARC) 69008 Lyon, France.

Dinand, V. & Arya, L. S. (2006). Epidemiology of childhood Hodgkins disease: is it different in developing countries? *Indian Pediatr* **43**, 141-147.

Dolan, A., Addison, C., Gatherer, D., Davison, A. J. & McGeoch, D. J. (2006). The genome of Epstein-Barr virus type 2 strain AG876. *Virology* **350**, 164-170.

Dolan, A., Cunningham, C., Hector, R. D., Hassan-Walker, A. F., Lee, L., Addison, C., Dargan, D. J., McGeoch, D. J., Gatherer, D., Emery, V. C., Griffiths, P. D., Sinzger, C., McSharry, B. P., Wilkinson, G. W. & Davison, A. J. (2004). Genetic content of wild-type human cytomegalovirus. *J Gen Virol* **85**, 1301-1312.

Donati, D., Mok, B., Chene, A., Xu, H., Thangarajh, M., Glas, R., Chen, Q., Wahlgren, M. & Bejarano, M. T. (2006). Increased B cell survival and preferential activation of the memory compartment by a malaria polyclonal B cell activator. *J Immunol* **177**, 3035-3044.

Dono, M., Burgio, V. L., Tacchetti, C., Favre, A., Augliera, A., Zupo, S., Taborelli, G., Chiorazzi, N., Grossi, C. E. & Ferrarini, M. (1996). Subepithelial B cells in the human palatine tonsil. I. Morphologic, cytochemical and phenotypic characterization. *Eur J Immunol* **26**, 2035-2042.

Dressman, D., Yan, H., Traverso, G., Kinzler, K. W. & Vogelstein, B. (2003). Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc Natl Acad Sci U S A* **100**, 8817-8822.

Driver, D. J., McHeyzer-Williams, L. J., Cool, M., Stetson, D. B. & McHeyzer-Williams, M. G. (2001). Development and maintenance of a B220- memory B cell compartment. *J Immunol* **167**, 1393-1405.

Dror, Y., Greenberg, M., Taylor, G., Superina, R., Hebert, D., West, L., Connolly, B., Sena, L., Allen, U. & Weitzman, S. (1999). Lymphoproliferative disorders after organ transplantation in children. *Transplantation* **67**, 990-998.

Drummond, A. J. & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214.

Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* **22**, 1185-1192.

DuBridge, R. B., Tang, P., Hsia, H. C., Leong, P. M., Miller, J. H. & Calos, M. P. (1987). Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. *Mol Cell Biol* **7**, 379-387.

Dyson, P. J. & Farrell, P. J. (1985). Chromatin structure of Epstein-Barr virus. *J Gen Virol* **66 (Pt 9)**, 1931-1940.

Edwards, R. H., Seillier-Moiseiwitsch, F. & Raab-Traub, N. (1999). Signature amino acid changes in latent membrane protein 1 distinguish Epstein-Barr virus strains. *Virology* **261**, 79-95.

Edwards, R. H., Sitki-Green, D., Moore, D. T. & Raab-Traub, N. (2004). Potential selection of LMP1 variants in nasopharyngeal carcinoma. *J Virol* **78**, 868-881.

Ehrhardt, G. R., Hsu, J. T., Gartland, L., Leu, C. M., Zhang, S., Davis, R. S. & Cooper, M. D. (2005). Expression of the immunoregulatory molecule FcRH4 defines a distinctive tissue-based population of memory B cells. *J Exp Med* **202**, 783-791.

Eliopoulos, A. G., Gallagher, N. J., Blake, S. M., Dawson, C. W. & Young, L. S. (1999). Activation of the p38 mitogen-activated protein kinase pathway by Epstein-Barr virus-encoded latent membrane protein 1 coregulates interleukin-6 and interleukin-8 production. *J Biol Chem* **274**, 16085-16096.

Epstein, M. A., Achong, B. G. & Barr, Y. M. (1964). Virus Particles in Cultured Lymphoblasts from Burkitt's Lymphoma. *Lancet* **1**, 702-703.

Fahmi, H., Cochet, C., Hmama, Z., Opolon, P. & Joab, I. (2000). Transforming growth factor beta 1 stimulates expression of the Epstein-Barr virus BZLF1 immediate-early gene product ZEBRA by an indirect mechanism which requires the MAPK kinase pathway. *J Virol* **74**, 5810-5818.

Fahraeus, R., Jansson, A., Ricksten, A., Sjoblom, A. & Rymo, L. (1990). Epstein-Barr virus-encoded nuclear antigen 2 activates the viral latent membrane protein promoter by modulating the activity of a negative regulatory element. *Proc Natl Acad Sci U S A* **87**, 7390-7394.

Fahraeus, R., Jansson, A., Sjoblom, A., Nilsson, T., Klein, G. & Rymo, L. (1993). Cell phenotype-dependent control of Epstein-Barr virus latent membrane protein 1 gene regulatory sequences. *Virology* **195**, 71-80.

Fairfax, K. A., Corcoran, L. M., Pridans, C., Huntington, N. D., Kallies, A., Nutt, S. L. & Tarlinton, D. M. (2007). Different kinetics of blimp-1 induction in B cell subsets revealed by reporter gene. *J Immunol* **178**, 4104-4111.

Fais, F., Gaidano, G., Capello, D., Gloghini, A., Ghiotto, F., Roncella, S., Carbone, A., Chiorazzi, N. & Ferrarini, M. (1999). Immunoglobulin V region gene use and structure suggest antigen selection in AIDS-related primary effusion lymphomas. *Leukemia* **13**, 1093-1099.

Fan, W., Bubman, D., Chadburn, A., Harrington, W. J., Jr., Cesarman, E. & Knowles, D. M. (2005). Distinct subsets of primary effusion lymphoma can be identified based on their cellular gene expression profile and viral association. *J Virol* **79**, 1244-1251.

Farrell, P. J., Rowe, D. T., Rooney, C. M. & Kouzarides, T. (1989). Epstein-Barr virus BZLF1 trans-activator specifically binds to a consensus AP-1 site and is related to c-fos. *EMBO J* **8**, 127-132.

Fearon, D. T., Manders, P. M. & Wagner, S. D. (2002). Bcl-6 uncouples B lymphocyte proliferation from differentiation. *Adv Exp Med Biol* **512**, 21-28.

Fedurco, M., Romieu, A., Williams, S., Lawrence, I. & Turcatti, G. (2006). BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res* **34**, e22.

Feederle, R., Kost, M., Baumann, M., Janz, A., Drouet, E., Hammerschmidt, W. & Delecluse, H. J. (2000). The Epstein-Barr virus lytic program is controlled by the co-operative functions of two transactivators. *EMBO J* **19**, 3080-3089.

Feng, W. H. & Kenney, S. C. (2006). Valproic acid enhances the efficacy of chemotherapy in EBV-positive tumors by increasing lytic viral gene expression. *Cancer Res* **66**, 8762-8769.

Feng, W. H., Kraus, R. J., Dickerson, S. J., Lim, H. J., Jones, R. J., Yu, X., Mertz, J. E. & Kenney, S. C. (2007). ZEB1 and c-Jun levels contribute to the establishment of highly lytic Epstein-Barr virus infection in gastric AGS cells. *J Virol* **81**, 10113-10122.

Ferlay, J. (1992). Cancer Incidence in Five Continents. Processing of data. *IARC Sci Publ*, 39-44.

Fingeroth, J. D., Clabby, M. L. & Strominger, J. D. (1988). Characterization of a T-lymphocyte Epstein-Barr virus/C3d receptor (CD21). *J Virol* **62**, 1442-1447.

Fingeroth, J. D., Weis, J. J., Tedder, T. F., Strominger, J. L., Biro, P. A. & Fearon, D. T. (1984). Epstein-Barr virus receptor of human B lymphocytes is the C3d receptor CR2. *Proc Natl Acad Sci U S A* **81**, 4510-4514.

Fixman, E. D., Hayward, G. S. & Hayward, S. D. (1992). trans-acting requirements for replication of Epstein-Barr virus ori-Lyt. *J Virol* **66**, 5030-5039.

Flemington, E. & Speck, S. H. (1990a). Autoregulation of Epstein-Barr virus putative lytic switch gene BZLF1. *J Virol* **64**, 1227-1232.

Flemington, E. & Speck, S. H. (1990b). Evidence for coiled-coil dimer formation by an Epstein-Barr virus transactivator that lacks a heptad repeat of leucine residues. *Proc Natl Acad Sci U S A* **87**, 9459-9463.

Flemington, E. & Speck, S. H. (1990c). Identification of phorbol ester response elements in the promoter of Epstein-Barr virus putative lytic switch gene BZLF1. *J Virol* **64**, 1217-1226.

Flower, K., Thomas, D., Heather, J., Ramasubramanyan, S., Jones, S. & Sinclair, A. J. (2011). Epigenetic control of viral life-cycle by a DNA-methylation dependent transcription factor. *PLoS One* **6**, e25922.

Flynn, G. C., Pohl, J., Flocco, M. T. & Rothman, J. E. (1991). Peptide-binding specificity of the molecular chaperone BiP. *Nature* **353**, 726-730.

Ford, P. W., Bryan, B. A., Dyson, O. F., Weidner, D. A., Chintalgattu, V. & Akula, S. M. (2006). Raf/MEK/ERK signalling triggers reactivation of Kaposi's sarcoma-associated herpesvirus latency. *J Gen Virol* **87**, 1139-1144.

Forman, S. J., Zaia, J. A., Clark, B. R., Wright, C. L., Mills, B. J., Pottathil, R., Racklin, B. C., Gallagher, M. T., Welte, K. & Blume, K. G. (1985). A 64,000 dalton matrix protein of human cytomegalovirus induces in vitro immune responses similar to those of whole viral antigen. *J Immunol* **134**, 3391-3395.

Foster, K. M. (1966). Post-transfusion mononucleosis. *Australas Ann Med* **15**, 305-310.

Frappier, L. & O'Donnell, M. (1991). Epstein-Barr nuclear antigen 1 mediates a DNA loop within the latent replication origin of Epstein-Barr virus. *Proc Natl Acad Sci U S A* **88**, 10875-10879.

Fruehling, S. & Longnecker, R. (1997). The immunoreceptor tyrosine-based activation motif of Epstein-Barr virus LMP2A is essential for blocking BCR-mediated signal transduction. *Virology* **235**, 241-251.

Fruscalzo, A., Marsili, G., Busiello, V., Bertolini, L. & Frezza, D. (2001). DNA sequence heterogeneity within the Epstein-Barr virus family of repeats in the latent origin of replication. *Gene* **265**, 165-173.

Furlong, D., Swift, H. & Roizman, B. (1972). Arrangement of herpesvirus deoxyribonucleic acid in the core. *J Virol* **10**, 1071-1074.

Furnari, F. B., Zacny, V., Quinlivan, E. B., Kenney, S. & Pagano, J. S. (1994). RAZ, an Epstein-Barr virus transdominant repressor that modulates the viral reactivation mechanism. *J Virol* **68**, 1827-1836.

Gaidano, G., Cechova, K., Chang, Y., Moore, P. S., Knowles, D. M. & Dalla-Favera, R. (1996). Establishment of AIDS-related lymphoma cell lines from lymphomatous effusions. *Leukemia* **10**, 1237-1240.

Gaidano, G., Gloghini, A., Gattei, V., Rossi, M. F., Cilia, A. M., Godeas, C., Degan, M., Perin, T., Canzonieri, V., Aldinucci, D., Saglio, G., Carbone, A. & Pinto, A. (1997). Association of Kaposi's sarcoma-associated herpesvirus-positive primary effusion lymphoma with expression of the CD138/syndecan-1 antigen. *Blood* **90**, 4894-4900.

Gao, Z., Krithivas, A., Finan, J. E., Semmes, O. J., Zhou, S., Wang, Y. & Hayward, S. D. (1998). The Epstein-Barr virus lytic transactivator Zta interacts with the helicase-primase replication proteins. *J Virol* **72**, 8559-8567.

Gardella, T., Medveczky, P., Sairenji, T. & Mulder, C. (1984). Detection of circular and linear herpesvirus DNA molecules in mammalian cells by gel electrophoresis. *J Virol* **50**, 248-254.

Gardner, B. M. & Walter, P. (2011). Unfolded proteins are Ire1-activating ligands that directly induce the unfolded protein response. *Science* **333**, 1891-1894.

Gass, J. N., Gifford, N. M. & Brewer, J. W. (2002). Activation of an unfolded protein response during differentiation of antibody-secreting B cells. *J Biol Chem* **277**, 49047-49054.

Gass, J. N., Gunn, K. E., Sriburi, R. & Brewer, J. W. (2004). Stressed-out B cells? Plasma-cell differentiation and the unfolded protein response. *Trends Immunol* **25**, 17-24.

Gerber, P., Lucas, S., Nonoyama, M., Perlin, E. & Goldstein, L. I. (1972). Oral excretion of Epstein-Barr virus by healthy subjects and patients with infectious mononucleosis. *Lancet* **2**, 988-989.

Gibbs, M. J., Armstrong, J. S. & Gibbs, A. J. (2000). Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* **16**, 573-582.

Glenn, M., Rainbow, L., Aurade, F., Davison, A. & Schulz, T. F. (1999). Identification of a spliced gene from Kaposi's sarcoma-associated herpesvirus encoding a protein with similarities to latent membrane proteins 1 and 2A of Epstein-Barr virus. *J Virol* **73**, 6953-6963.

Good, K. L., Avery, D. T. & Tangye, S. G. (2009). Resting human memory B cells are intrinsically programmed for enhanced survival and responsiveness to diverse stimuli compared to naive B cells. *J Immunol* **182**, 890-901.

Goodnow, C. C., Crosbie, J., Adelstein, S., Lavoie, T. B., Smith-Gill, S. J., Mason, D. Y., Jorgensen, H., Brink, R. A., Pritchard-Briscoe, H., Loughnan, M. & et al. (1989). Clonal silencing of self-reactive B lymphocytes in a transgenic mouse model. *Cold Spring Harb Symp Quant Biol* **54 Pt 2**, 907-920.

Goossens, T., Klein, U. & Kuppers, R. (1998). Frequent occurrence of deletions and duplications during somatic hypermutation: implications for oncogene translocations and heavy chain disease. *Proc Natl Acad Sci U S A* **95**, 2463-2468.

Gorzer, I., Niesters, H. G., Cornelissen, J. J. & Puchhammer-Stockl, E. (2006). Characterization of Epstein-Barr virus Type I variants based on linked polymorphism among EBNA3A, -3B, and -3C genes. *Virus Res* **118**, 105-114.

Gottschalk, S., Ng, C. Y., Perez, M., Smith, C. A., Sample, C., Brenner, M. K., Heslop, H. E. & Rooney, C. M. (2001). An Epstein-Barr virus deletion mutant associated with fatal lymphoproliferative disease unresponsive to therapy with virus-specific CTLs. *Blood* **97**, 835-843.

Gradoville, L., Gerlach, J., Grogan, E., Shedd, D., Nikiforow, S., Metroka, C. & Miller, G. (2000). Kaposi's sarcoma-associated herpesvirus open reading frame 50/Rta protein activates the entire viral lytic cycle in the HH-B2 primary effusion lymphoma cell line. *J Virol* **74**, 6207-6212.

Gradoville, L., Kwa, D., El-Guindy, A. & Miller, G. (2002). Protein kinase C-independent activation of the Epstein-Barr virus lytic cycle. *J Virol* **76**, 5612-5626.

Gregory, C. D., Rowe, M. & Rickinson, A. B. (1990). Different Epstein-Barr virus-B cell interactions in phenotypically distinct clones of a Burkitt's lymphoma cell line. *J Gen Virol* **71 (Pt 7)**, 1481-1495.

Gregory, C. D., Tursz, T., Edwards, C. F., Tetaud, C., Talbot, M., Caillou, B., Rickinson, A. B. & Lipinski, M. (1987). Identification of a subset of normal B cells with a Burkitt's lymphoma (BL)-like phenotype. *J Immunol* **139**, 313-318.

Grossman, S. R., Johannsen, E., Tong, X., Yalamanchili, R. & Kieff, E. (1994). The Epstein-Barr virus nuclear antigen 2 transactivator is directed to response elements by the J kappa recombination signal binding protein. *Proc Natl Acad Sci U S A* **91**, 7568-7572.

Groves, A. K., Cotter, M. A., Subramanian, C. & Robertson, E. S. (2001). The latency-associated nuclear antigen encoded by Kaposi's sarcoma-associated herpesvirus activates two major essential Epstein-Barr virus latent promoters. *J Virol* **75**, 9446-9457.

Gruffat, H., Manet, E. & Sergeant, A. (2002). MEF2-mediated recruitment of class II HDAC at the EBV immediate early gene BZLF1 links latency and chromatin remodeling. *EMBO Rep* **3**, 141-146.

Guasparri, I., Bubman, D. & Cesarman, E. (2008). EBV LMP2A affects LMP1-mediated NF-kappaB signaling and survival of lymphoma cells by regulating TRAF2 expression. *Blood* **111**, 3813-3820.

Gunn, M. D., Ngo, V. N., Ansel, K. M., Ekland, E. H., Cyster, J. G. & Williams, L. T. (1998). A B-cell-homing chemokine made in lymphoid follicles activates Burkitt's lymphoma receptor-1. *Nature* **391**, 799-803.

Guthridge, J. M., Young, K., Gipson, M. G., Sarrias, M. R., Szakonyi, G., Chen, X. S., Malaspina, A., Donoghue, E., James, J. A., Lambris, J. D., Moir, S. A., Perkins, S. J. & Holers, V. M. (2001). Epitope mapping using the X-ray crystallographic structure of complement receptor type 2 (CR2)/CD21: identification of a highly inhibitory monoclonal antibody that directly recognizes the CR2-C3d interface. *J Immunol* **167**, 5758-5766.

Gutierrez, M. I., Bhatia, K., Barriga, F., Diez, B., Muriel, F. S., de Andreas, M. L., Epelman, S., Risueno, C. & Magrath, I. T. (1992). Molecular epidemiology of Burkitt's lymphoma from South America: differences in breakpoint location and Epstein-Barr virus association from tumors in other world regions. *Blood* **79**, 3261-3266.

Gutierrez, M. I., Ibrahim, M. M., Dale, J. K., Greiner, T. C., Straus, S. E. & Bhatia, K. (2002). Discrete alterations in the BZLF1 promoter in tumor and non-tumor-associated Epstein-Barr virus. *J Natl Cancer Inst* **94**, 1757-1763.

Gutierrez, M. I., Raj, A., Spangler, G., Sharma, A., Hussain, A., Judde, J. G., Tsao, S. W., Yuen, P. W., Joab, I., Magrath, I. T. & Bhatia, K. (1997). Sequence variations in EBNA-1 may dictate restriction of tissue distribution of Epstein-Barr virus in normal and tumour cells. *J Gen Virol* **78 (Pt 7)**, 1663-1670.

Haan, K. M., Lee, S. K. & Longnecker, R. (2001). Different functional domains in the cytoplasmic tail of glycoprotein B are involved in Epstein-Barr virus-induced membrane fusion. *Virology* **290**, 106-114.

Haas, I. G. & Wabl, M. (1983). Immunoglobulin heavy chain binding protein. *Nature* **306**, 387-389.

Habeshaw, G., Yao, Q. Y., Bell, A. I., Morton, D. & Rickinson, A. B. (1999). Epstein-barr virus nuclear antigen 1 sequences in endemic and sporadic Burkitt's lymphoma reflect virus strains prevalent in different geographic areas. *J Virol* **73**, 965-975.

Hackney, J. A., Misaghi, S., Senger, K., Garris, C., Sun, Y., Lorenzo, M. N. & Zarrin, A. A. (2009). DNA targets of AID evolutionary link between antibody somatic hypermutation and class switch recombination. *Adv Immunol* **101**, 163-189.

Hadinoto, V., Shapiro, M., Sun, C. C. & Thorley-Lawson, D. A. (2009). The dynamics of EBV shedding implicate a central role for epithelial cells in amplifying viral output. *PLoS Pathog* **5**, e1000496.

Halder, S., Murakami, M., Verma, S. C., Kumar, P., Yi, F. & Robertson, E. S. (2009). Early events associated with infection of Epstein-Barr virus infection of primary B-cells. *PLoS One* **4**, e7214.

Hammerschmidt, W. & Sugden, B. (1988). Identification and characterization of oriLyt, a lytic origin of DNA replication of Epstein-Barr virus. *Cell* **55**, 427-433.

Harada, S. & Kieff, E. (1997). Epstein-Barr virus nuclear protein LP stimulates EBNA-2 acidic domain-mediated transcriptional activation. *J Virol* **71**, 6611-6618.

Harding, H. P., Zhang, Y., Bertolotti, A., Zeng, H. & Ron, D. (2000). Perk is essential for translational regulation and cell survival during the unfolded protein response. *Mol Cell* **5**, 897-904.

Harding, H. P., Zhang, Y. & Ron, D. (1999). Protein translation and folding are coupled by an endoplasmic-reticulum-resident kinase. *Nature* **397**, 271-274.

Harding, H. P., Zhang, Y., Zeng, H., Novoa, I., Lu, P. D., Calfon, M., Sadri, N., Yun, C., Popko, B., Paules, R., Stojdl, D. F., Bell, J. C., Hettmann, T., Leiden, J. M. & Ron, D. (2003). An integrated stress response regulates amino acid metabolism and resistance to oxidative stress. *Mol Cell* **11**, 619-633.

Hardy, R. R. & Hayakawa, K. (1991). A developmental switch in B lymphopoiesis. *Proc Natl Acad Sci U S A* **88**, 11550-11554.

Harmon, W. E. & Dharnidharka, V. R. (1999). Lymphoproliferative disease in children. *Transplant Proc* **31**, 1268-1269.

Harris, N. L., Ferry, J. A. & Swerdlow, S. H. (1997). Posttransplant lymphoproliferative disorders: summary of Society for Hematopathology Workshop. *Semin Diagn Pathol* **14**, 8-14.

Harris, N. L., Swerdlow, S. H., Frizzera, G. & Knowles, D. M. (2001). Post-transplant lymphoproliferative disorders. In *Pathology and Genetics of Tumour of Haematopoietic and Lymphoid tissues*. Edited by E. S. Jaffe, N. L. Harris, H. Stein & J. W. Vardiman. Lyon: IARC press.

Harwood, N. E. & Batista, F. D. (2008). New insights into the early molecular events underlying B cell activation. *Immunity* **28**, 609-619.

Hayward, G. S. (1999). KSHV strains: the origins and global spread of the virus. *Semin Cancer Biol* **9**, 187-199.

Hayward, G. S. & Zong, J. C. (2007). Modern evolutionary history of the human KSHV genome. *Curr Top Microbiol Immunol* **312**, 1-42.

Haze, K., Yoshida, H., Yanagi, H., Yura, T. & Mori, K. (1999). Mammalian transcription factor ATF6 is synthesized as a transmembrane protein and activated by proteolysis in response to endoplasmic reticulum stress. *Mol Biol Cell* **10**, 3787-3799.

Henderson, S., Huen, D., Rowe, M., Dawson, C., Johnson, G. & Rickinson, A. (1993). Epstein-Barr virus-coded BHRF1 protein, a viral homologue of Bcl-2, protects human B cells from programmed cell death. *Proc Natl Acad Sci U S A* **90**, 8479-8483.

Henderson, S., Rowe, M., Gregory, C., Croom-Carter, D., Wang, F., Longnecker, R., Kieff, E. & Rickinson, A. (1991). Induction of bcl-2 expression by Epstein-Barr virus latent membrane protein 1 protects infected B cells from programmed cell death. *Cell* **65**, 1107-1115.

Henkel, T., Ling, P. D., Hayward, S. D. & Peterson, M. G. (1994). Mediation of Epstein-Barr virus EBNA2 transactivation by recombination signal-binding protein J kappa. *Science* **265**, 92-95.

Hetz, C., Bernasconi, P., Fisher, J., Lee, A. H., Bassik, M. C., Antonsson, B., Brandt, G. S., Iwakoshi, N. N., Schinzel, A., Glimcher, L. H. & Korsmeyer, S. J. (2006). Proapoptotic BAX and BAK modulate the unfolded protein response by a direct interaction with IRE1alpha. *Science* **312**, 572-576.

Ho, M., Jaffe, R., Miller, G., Breinig, M. K., Dummer, J. S., Makowka, L., Atchison, R. W., Karrer, F., Nalesnik, M. A. & Starzl, T. E. (1988). The frequency of

Epstein-Barr virus infection and associated lymphoproliferative syndrome after transplantation and its manifestations in children. *Transplantation* **45**, 719-727.

Hochberg, D., Middeldorp, J. M., Catalina, M., Sullivan, J. L., Luzuriaga, K. & Thorley-Lawson, D. A. (2004). Demonstration of the Burkitt's lymphoma Epstein-Barr virus phenotype in dividing latently infected memory cells in vivo. *Proc Natl Acad Sci U S A* **101**, 239-244.

Hong, M., Luo, S., Baumeister, P., Huang, J. M., Gogia, R. K., Li, M. & Lee, A. S. (2004). Underglycosylation of ATF6 as a novel sensing mechanism for activation of the unfolded protein response. *J Biol Chem* **279**, 11354-11363.

Hosoi, T., Kume, A., Otani, K., Oba, T. & Ozawa, K. (2010). A unique modulator of endoplasmic reticulum stress-signalling pathways: the novel pharmacological properties of amiloride in glial cells. *Br J Pharmacol* **159**, 428-437.

Hsu, J. L. & Glaser, S. L. (2000). Epstein-barr virus-associated malignancies: epidemiologic patterns and etiologic implications. *Crit Rev Oncol Hematol* **34**, 27-53.

Hu, C. C., Dougan, S. K., McGehee, A. M., Love, J. C. & Ploegh, H. L. (2009). XBP-1 regulates signal transduction, transcription factors and bone marrow colonization in B cells. *EMBO J* **28**, 1624-1636.

Hummel, M., Thorley-Lawson, D. & Kieff, E. (1984). An Epstein-Barr virus DNA fragment encodes messages for the two major envelope glycoproteins (gp350/300 and gp220/200). *J Virol* **49**, 413-417.

Hutt-Fletcher, L. M. (2007). Epstein-Barr virus entry. *J Virol* **81**, 7825-7832.

Itakura, O., Yamada, S., Narita, M. & Kikuta, H. (1996). High prevalence of a 30-base pair deletion and single-base mutations within the carboxy terminal end of the LMP-1 oncogene of Epstein-Barr virus in the Japanese population. *Oncogene* **13**, 1549-1553.

Iwakoshi, N. N., Lee, A. H., Vallabhajosyula, P., Otipoby, K. L., Rajewsky, K. & Glimcher, L. H. (2003). Plasma cell differentiation and the unfolded protein response intersect at the transcription factor XBP-1. *Nat Immunol* **4**, 321-329.

Jacob, J., Kelsoe, G., Rajewsky, K. & Weiss, U. (1991). Intraclonal generation of antibody mutants in germinal centres. *Nature* **354**, 389-392.

Jacquot, S., Kobata, T., Iwata, S., Morimoto, C. & Schlossman, S. F. (1997). CD154/CD40 and CD70/CD27 interactions have different and sequential functions in T cell-dependent B cell responses: enhancement of plasma cell differentiation by CD27 signaling. *J Immunol* **159**, 2652-2657.

James, J. A., Kaufman, K. M., Farris, A. D., Taylor-Albert, E., Lehman, T. J. & Harley, J. B. (1997). An increased prevalence of Epstein-Barr virus infection in young patients suggests a possible etiology for systemic lupus erythematosus. *J Clin Invest* **100**, 3019-3026.

Janeway, C. A., Jr., Travers, P., Walport, M. & Shlomchik, M. J. (2001). *Immunobiology*. New York: Garland Science.

Jenner, R. G., Maillard, K., Cattini, N., Weiss, R. A., Boshoff, C., Wooster, R. & Kellam, P. (2003). Kaposi's sarcoma-associated herpesvirus-infected primary effusion lymphoma has a plasma cell gene expression profile. *Proc Natl Acad Sci U S A* **100**, 10399-10404.

Ji, K. M., Li, C. L., Meng, G., Han, A. D. & Wu, X. L. (2008). New BZLF1 sequence variations in EBV-associated undifferentiated nasopharyngeal carcinoma in southern China. *Arch Virol* **153**, 1949-1953.

Jiang, Y., Xu, D., Zhao, Y. & Zhang, L. (2008). Mutual inhibition between Kaposi's sarcoma-associated herpesvirus and Epstein-Barr virus lytic replication initiators in dually-infected primary effusion lymphoma. *PLoS One* **3**, e1569.

Jin, Y., Xie, Z., Lu, G., Yang, S. & Shen, K. (2010). Characterization of variants in the promoter of BZLF1 gene of EBV in nonmalignant EBV-associated diseases in Chinese children. *Virol J* **7**, 92.

Johannsen, E., Koh, E., Mosialos, G., Tong, X., Kieff, E. & Grossman, S. R. (1995). Epstein-Barr virus nuclear protein 2 transactivation of the latent membrane protein 1 promoter is mediated by J kappa and PU.1. *J Virol* **69**, 253-262.

Jones, C. H., Hayward, S. D. & Rawlins, D. R. (1989). Interaction of the lymphocyte-derived Epstein-Barr virus nuclear antigen EBNA-1 with its DNA-binding sites. *J Virol* **63**, 101-110.

Joseph, A. M., Babcock, G. J. & Thorley-Lawson, D. A. (2000). Cells expressing the Epstein-Barr virus growth program are present in and restricted to the naive B-cell subset of healthy tonsils. *J Virol* **74**, 9964-9971.

Kafuko, G. W. & Burkitt, D. P. (1970). Burkitt's lymphoma and malaria. *Int J Cancer* **6**, 1-9.

Kanda, T., Otter, M. & Wahl, G. M. (2001). Coupling of mitotic chromosome tethering and replication competence in epstein-barr virus-based plasmids. *Mol Cell Biol* **21**, 3576-3588.

Kapoor, P., Lavoie, B. D. & Frappier, L. (2005). EBP2 plays a key role in Epstein-Barr virus mitotic segregation and is regulated by aurora family kinases. *Mol Cell Biol* **25**, 4934-4945.

Kaposi, M. (1872). Idiopathic multiple pigmented sarcoma of the skin. *Arch Dermatol Syphil* **4**, 265--273.

Karasuyama, H., Kudo, A. & Melchers, F. (1990). The proteins encoded by the VpreB and lambda 5 pre-B cell-specific genes can associate with each other and with mu heavy chain. *J Exp Med* **172**, 969-972.

Karlsson, Q. H., Schelcher, C., Verrall, E., Petosa, C. & Sinclair, A. J. (2008a). Methylated DNA recognition during the reversal of epigenetic silencing is regulated by cysteine and serine residues in the Epstein-Barr virus lytic switch protein. *PLoS Pathog* **4**, e1000005.

Karlsson, Q. H., Schelcher, C., Verrall, E., Petosa, C. & Sinclair, A. J. (2008b). The reversal of epigenetic silencing of the EBV genome is regulated by viral bZIP protein. *Biochem Soc Trans* **36**, 637-639.

Kaye, K. M., Izumi, K. M. & Kieff, E. (1993). Epstein-Barr virus latent membrane protein 1 is essential for B-lymphocyte growth transformation. *Proc Natl Acad Sci U S A* **90**, 9150-9154.

Keller, A. D. & Maniatis, T. (1991). Identification and characterization of a novel repressor of beta-interferon gene expression. *Genes Dev* **5**, 868-879.

Kelly, G., Bell, A. & Rickinson, A. (2002). Epstein-Barr virus-associated Burkitt lymphomagenesis selects for downregulation of the nuclear antigen EBNA2. *Nat Med* **8**, 1098-1104.

Kelly, G. L., Long, H. M., Stylianou, J., Thomas, W. A., Leese, A., Bell, A. I., Bornkamm, G. W., Mautner, J., Rickinson, A. B. & Rowe, M. (2009). An Epstein-Barr virus anti-apoptotic protein constitutively expressed in transformed cells and implicated in burkitt lymphomagenesis: the Wp/BHRF1 link. *PLoS Pathog* **5**, e1000341.

Kelly, G. L., Milner, A. E., Baldwin, G. S., Bell, A. I. & Rickinson, A. B. (2006). Three restricted forms of Epstein-Barr virus latency counteracting

apoptosis in c-myc-expressing Burkitt lymphoma cells. *Proc Natl Acad Sci U S A* **103**, 14935-14940.

Kelly, G. L., Milner, A. E., Tierney, R. J., Croom-Carter, D. S., Altmann, M., Hammerschmidt, W., Bell, A. I. & Rickinson, A. B. (2005). Epstein-Barr virus nuclear antigen 2 (EBNA2) gene deletion is consistently linked with EBNA3A, -3B, and -3C expression in Burkitt's lymphoma cells and with increased resistance to apoptosis. *J Virol* **79**, 10709-10717.

Kenney, S. C., Holley-Guthrie, E., Quinlivan, E. B., Gutsch, D., Zhang, Q., Bender, T., Giot, J. F. & Sergeant, A. (1992). The cellular oncogene c-myb can interact synergistically with the Epstein-Barr virus BZLF1 transactivator in lymphoid cells. *Mol Cell Biol* **12**, 136-146.

Kieser, A., Kilger, E., Gires, O., Ueffing, M., Kolch, W. & Hammerschmidt, W. (1997). Epstein-Barr virus latent membrane protein-1 triggers AP-1 activity *via* the c-Jun N-terminal kinase cascade. *EMBO J* **16**, 6478-6485.

Kilger, E., Kieser, A., Baumann, M. & Hammerschmidt, W. (1998). Epstein-Barr virus-mediated B-cell proliferation is dependent upon latent membrane protein 1, which simulates an activated CD40 receptor. *EMBO J* **17**, 1700-1709.

Kimata, Y., Ishiwata-Kimata, Y., Ito, T., Hirata, A., Suzuki, T., Oikawa, D., Takeuchi, M. & Kohno, K. (2007). Two regulatory steps of ER-stress sensor Ire1 involving its cluster formation and interaction with unfolded proteins. *J Cell Biol* **179**, 75-86.

Kingma, D. W., Weiss, W. B., Jaffe, E. S., Kumar, S., Frekko, K. & Raffeld, M. (1996). Epstein-Barr virus latent membrane protein-1 oncogene deletions: correlations with malignancy in Epstein-Barr virus--associated lymphoproliferative disorders and malignant lymphomas. *Blood* **88**, 242-251.

Kirchmaier, A. L. & Sugden, B. (1995). Plasmid maintenance of derivatives of oriP of Epstein-Barr virus. *J Virol* **69**, 1280-1283.

Klee, M., Pallauf, K., Alcala, S., Fleischer, A. & Pimentel-Muinos, F. X. (2009). Mitochondrial apoptosis induced by BH3-only molecules in the exclusive presence of endoplasmic reticular Bak. *EMBO J* **28**, 1757-1768.

Klein, E., Klein, G., Nadkarni, J. S., Nadkarni, J. J., Wigzell, H. & Clifford, P. (1967). Surgace IgM specificity on cells derived from a Burkitt's lymphoma. *Lancet* **2**, 1068-1070.

Klein, E., Klein, G., Nadkarni, J. S., Nadkarni, J. J., Wigzell, H. & Clifford, P. (1968). Surface IgM-kappa specificity on a Burkitt lymphoma cell in vivo and in derived culture lines. *Cancer Res* **28**, 1300-1310.

Klein, G. (1983). Specific chromosomal translocations and the genesis of B-cell-derived tumors in mice and men. *Cell* **32**, 311-315.

Klein, G., Klein, E. & Kashuba, E. (2010). Interaction of Epstein-Barr virus (EBV) with human B-lymphocytes. *Biochem Biophys Res Commun* **396**, 67-73.

Klein, R. & Zachau, H. G. (1995). Expression and hypermutation of human immunoglobulin kappa genes. *Ann N Y Acad Sci* **764**, 74-83.

Klein, U. & Dalla-Favera, R. (2008). Germinal centres: role in B-cell physiology and malignancy. *Nat Rev Immunol* **8**, 22-33.

Klein, U., Gloghini, A., Gaidano, G., Chadburn, A., Cesarman, E., Dalla-Favera, R. & Carbone, A. (2003). Gene expression profile analysis of AIDS-related primary effusion lymphoma (PEL) suggests a plasmablastic derivation and identifies PEL-specific transcripts. *Blood* **101**, 4115-4121.

Klein, U., Rajewsky, K. & Kuppers, R. (1998). Human immunoglobulin (Ig)M+IgD+ peripheral blood B cells expressing the CD27 cell surface antigen carry somatically mutated variable region genes: CD27 as a general marker for somatically mutated (memory) B cells. *J Exp Med* **188**, 1679-1689.

Klose, R. J. & Bird, A. P. (2006). Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* **31**, 89-97.

Klumb, C. E., Hassan, R., De Oliveira, D. E., De Resende, L. M., Carrico, M. K., De Almeida Dobbin, J., Pombo-De-Oliveira, M. S., Bacchi, C. E. & Maia, R. C. (2004). Geographic variation in Epstein-Barr virus-associated Burkitt's lymphoma in children from Brazil. *Int J Cancer* **108**, 66-70.

Knecht, H., Bachmann, E., Brousset, P., Sandvej, K., Nadal, D., Bachmann, F., Odermatt, B. F., Delsol, G. & Pallesen, G. (1993). Deletions within the LMP1 oncogene of Epstein-Barr virus are clustered in Hodgkin's disease and identical to those observed in nasopharyngeal carcinoma. *Blood* **82**, 2937-2942.

Knight, J. S., Lan, K., Subramanian, C. & Robertson, E. S. (2003). Epstein-Barr virus nuclear antigen 3C recruits histone deacetylase activity and associates with the corepressors mSin3A and NCoR in human B-cell lines. *J Virol* **77**, 4261-4272.

Knight, J. S., Sharma, N. & Robertson, E. S. (2005a). Epstein-Barr virus latent antigen 3C can mediate the degradation of the retinoblastoma protein through an SCF cellular ubiquitin ligase. *Proc Natl Acad Sci U S A* **102**, 18562-18566.

Knight, J. S., Sharma, N. & Robertson, E. S. (2005b). SCFSkp2 complex targeted by Epstein-Barr virus essential nuclear antigen. *Mol Cell Biol* **25**, 1749-1763.

Knowles, D. M. (2003). Etiology and pathogenesis of AIDS-related non-Hodgkin's lymphoma. *Hematol Oncol Clin North Am* **17**, 785-820.

Knowles, D. M., Inghirami, G., Ubriaco, A. & Dalla-Favera, R. (1989). Molecular genetic analysis of three AIDS-associated neoplasms of uncertain lineage demonstrates their B-cell derivation and the possible pathogenetic role of the Epstein-Barr virus. *Blood* **73**, 792-799.

Koboldt, D. C., Chen, K., Wylie, T., Larson, D. E., McLellan, M. D., Mardis, E. R., Weinstock, G. M., Wilson, R. K. & Ding, L. (2009). VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* **25**, 2283-2285.

Komanduri, K. V., Luce, J. A., McGrath, M. S., Herndier, B. G. & Ng, V. L. (1996). The natural history and molecular heterogeneity of HIV-associated primary malignant lymphomatous effusions. *J Acquir Immune Defic Syndr Hum Retrovirol* **13**, 215-226.

Komori, T., Okada, A., Stewart, V. & Alt, F. W. (1993). Lack of N regions in antigen receptor variable region genes of TdT-deficient lymphocytes. *Science* **261**, 1171-1175.

Kouzarides, T., Packham, G., Cook, A. & Farrell, P. J. (1991). The BZLF1 protein of EBV has a coiled coil dimerisation domain without a heptad leucine repeat but with homology to the C/EBP leucine zipper. *Oncogene* **6**, 195-204.

Kozmik, Z., Wang, S., Dorfler, P., Adams, B. & Busslinger, M. (1992). The promoter of the CD19 gene is a target for the B-cell-specific transcription factor BSAP. *Mol Cell Biol* **12**, 2662-2672.

Kraus, R. J., Mirocha, S. J., Stephany, H. M., Puchalski, J. R. & Mertz, J. E. (2001). Identification of a novel element involved in regulation of the lytic switch BZLF1 gene promoter of Epstein-Barr virus. *J Virol* **75**, 867-877.

Kraus, R. J., Perrigoue, J. G. & Mertz, J. E. (2003). ZEB negatively regulates the lytic-switch BZLF1 gene promoter of Epstein-Barr virus. *J Virol* **77**, 199-207.

Krithivas, A., Young, D. B., Liao, G., Greene, D. & Hayward, S. D. (2000). Human herpesvirus 8 LANA interacts with proteins of the mSin3 corepressor complex and negatively regulates Epstein-Barr virus gene expression in dually infected PEL cells. *J Virol* **74**, 9637-9645.

Kuo, T. C., Shaffer, A. L., Haddad, J., Jr., Choi, Y. S., Staudt, L. M. & Calame, K. (2007). Repression of BCL-6 is required for the formation of human memory B cells in vitro. *J Exp Med* **204**, 819-830.

Kuppers, R. (2003). B cells under influence: transformation of B cells by Epstein-Barr virus. *Nat Rev Immunol* **3**, 801-812.

Kurth, J., Hansmann, M. L., Rajewsky, K. & Kuppers, R. (2003). Epstein-Barr virus-infected B cells expanding in germinal centers of infectious mononucleosis patients do not participate in the germinal center reaction. *Proc Natl Acad Sci U S A* **100**, 4730-4735.

Lagunoff, M. & Ganem, D. (1997). The structure and coding organization of the genomic termini of Kaposi's sarcoma-associated herpesvirus. *Virology* **236**, 147-154.

Lai, I. Y., Farrell, P. J. & Kellam, P. (2011). X-box binding protein 1 induces the expression of the lytic cycle transactivator of Kaposi's sarcoma-associated herpesvirus but not Epstein-Barr virus in co-infected primary effusion lymphoma. *J Gen Virol* **92**, 421-431.

Laichalk, L. L. & Thorley-Lawson, D. A. (2005). Terminal differentiation into plasma cells initiates the replicative cycle of Epstein-Barr virus in vivo. *J Virol* **79**, 1296-1307.

Lanzavecchia, A. (1985). Antigen-specific interaction between T and B cells. *Nature* **314**, 537-539.

Le Roux, A., Kerdiles, B., Walls, D., Dedieu, J. F. & Perricaudet, M. (1994). The Epstein-Barr virus determined nuclear antigens EBNA-3A, -3B, and -3C repress EBNA-2-mediated transactivation of the viral terminal protein 1 gene promoter. *Virology* **205**, 596-602.

Leder, P. (1985). The state of and prospects for molecular genetics in Burkitt's lymphoma. *IARC Sci Publ*, 475-476.

Lee, A. H., Iwakoshi, N. N. & Glimcher, L. H. (2003). XBP-1 regulates a subset of endoplasmic reticulum resident chaperone genes in the unfolded protein response. *Mol Cell Biol* **23**, 7448-7459.

Lee, G., Namen, A. E., Gillis, S. & Kincade, P. W. (1988). Recombinant interleukin-7 supports the growth of normal B lymphocyte precursors. *Curr Top Microbiol Immunol* **141**, 16-18.

Lee, S. K. & Longnecker, R. (1997). The Epstein-Barr virus glycoprotein 110 carboxy-terminal tail domain is essential for lytic virus replication. *J Virol* **71**, 4092-4097.

Levine, M. H., Haberman, A. M., Sant'Angelo, D. B., Hannum, L. G., Cancro, M. P., Janeway, C. A., Jr. & Shlomchik, M. J. (2000). A B-cell receptor-specific selection step governs immature to mature B cell differentiation. *Proc Natl Acad Sci U S A* **97**, 2743-2748.

Levitskaya, J., Coram, M., Levitsky, V., Imreh, S., Steigerwald-Mullen, P. M., Klein, G., Kurilla, M. G. & Masucci, M. G. (1995). Inhibition of antigen

processing by the internal repeat region of the Epstein-Barr virus nuclear antigen-1. *Nature* **375**, 685-688.

Li, H. & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079.

Li, H. & Minarovits, J. (2003). Host cell-dependent expression of latent Epstein-Barr virus genomes: regulation by DNA methylation. *Adv Cancer Res* **89**, 133-156.

Li, Q., Spriggs, M. K., Kovats, S., Turk, S. M., Comeau, M. R., Nepom, B. & Hutt-Fletcher, L. M. (1997). Epstein-Barr virus uses HLA class II as a cofactor for infection of B lymphocytes. *J Virol* **71**, 4657-4662.

Li, Q., Turk, S. M. & Hutt-Fletcher, L. M. (1995). The Epstein-Barr virus (EBV) BZLF2 gene product associates with the gH and gL homologs of EBV and carries an epitope critical to infection of B cells but not of epithelial cells. *J Virol* **69**, 3987-3994.

Li, Z., Van Calcar, S., Qu, C., Cavenee, W. K., Zhang, M. Q. & Ren, B. (2003). A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. *Proc Natl Acad Sci U S A* **100**, 8164-8169.

Liang, C. L., Chen, J. L., Hsu, Y. P., Ou, J. T. & Chang, Y. S. (2002). Epstein-Barr virus BZLF1 gene is activated by transforming growth factor-beta through cooperativity of Smads and c-Jun/c-Fos proteins. *J Biol Chem* **277**, 23345-23357.

Liang, X., Collins, C. M., Mendel, J. B., Iwakoshi, N. N. & Speck, S. H. (2009). Gammaherpesvirus-driven plasma cell differentiation regulates virus reactivation from latently infected B lymphocytes. *PLoS Pathog* **5**, e1000677.

Lieberman, P. M. & Berk, A. J. (1990). In vitro transcriptional activation, dimerization, and DNA-binding specificity of the Epstein-Barr virus Zta protein. *J Virol* **64**, 2560-2568.

Lin, K. I., Angelin-Duclos, C., Kuo, T. C. & Calame, K. (2002). Blimp-1-dependent repression of Pax-5 is required for differentiation of B cells to immunoglobulin M-secreting plasma cells. *Mol Cell Biol* **22**, 4771-4780.

Lindahl, T., Adams, A., Bjursell, G., Bornkamm, G. W., Kaschka-Dierich, C. & Jehn, U. (1976). Covalently closed circular duplex DNA of Epstein-Barr virus in a human lymphoid cell line. *J Mol Biol* **102**, 511-530.

Liou, H. C., Boothby, M. R., Finn, P. W., Davidon, R., Nabavi, N., Zeleznik-Le, N. J., Ting, J. P. & Glimcher, L. H. (1990). A new member of the leucine zipper class of proteins that binds to the HLA DR alpha promoter. *Science* **247**, 1581-1584.

Liu, P., Fang, X., Feng, Z., Guo, Y. M., Peng, R. J., Liu, T., Huang, Z., Feng, Y., Sun, X., Xiong, Z., Guo, X., Pang, S. S., Wang, B., Lv, X., Feng, F. T., Li, D. J., Chen, L. Z., Feng, Q. S., Huang, W. L., Zeng, M. S., Bei, J. X., Zhang, Y. & Zeng, Y. X. (2011). Direct sequencing and characterization of a clinical isolate of Epstein-Barr virus from nasopharyngeal carcinoma tissue by using next-generation sequencing technology. *J Virol* **85**, 11291-11299.

Liu, P., Liu, S. & Speck, S. H. (1998). Identification of a negative cis element within the ZII domain of the Epstein-Barr virus lytic switch BZLF1 gene promoter. *J Virol* **72**, 8230-8239.

Liu, S., Borras, A. M., Liu, P., Suske, G. & Speck, S. H. (1997a). Binding of the ubiquitous cellular transcription factors Sp1 and Sp3 to the ZI domains in the Epstein-Barr virus lytic switch BZLF1 gene promoter. *Virology* **228**, 11-18.

Liu, S., Liu, P., Borras, A., Chatila, T. & Speck, S. H. (1997b). Cyclosporin A-sensitive induction of the Epstein-Barr virus lytic switch is mediated *via* a novel pathway involving a MEF2 family member. *EMBO J* **16**, 143-153.

Liu, Y. J., Oldfield, S. & MacLennan, I. C. (1988). Memory B cells in T cell-dependent antibody responses colonize the splenic marginal zones. *Eur J Immunol* **18**, 355-362.

Lo, K. W. & Huang, D. P. (2002). Genetic and epigenetic changes in nasopharyngeal carcinoma. *Semin Cancer Biol* **12**, 451-462.

Loder, F., Mutschler, B., Ray, R. J., Paige, C. J., Sideras, P., Torres, R., Lamers, M. C. & Carsetti, R. (1999). B cell development in the spleen takes place in discrete steps and is determined by the quality of B cell receptor-derived signals. *J Exp Med* **190**, 75-89.

Longerich, S., Basu, U., Alt, F. & Storb, U. (2006). AID in somatic hypermutation and class switch recombination. *Curr Opin Immunol* **18**, 164-174.

Longnecker, R. (2000). Epstein-Barr virus latency: LMP2, a regulator or means for Epstein-Barr virus persistence? *Adv Cancer Res* **79**, 175-200.

Longnecker, R. H.-F., L.; Jardetzky, T. (2009). Epstein-Barr Virus Entry. In *DNA Tumor Viruses*, pp. 355-378: Springer.

Lopez, R., Urquiza, M., Patino, H., Suarez, J., Reyes, C., Patarroyo, M. A. & Patarroyo, M. E. (2005). A B-lymphocyte binding peptide from BNRF1 induced antibodies inhibiting EBV-invasion of B-lymphocytes. *Biochimie* **87**, 985-992.

Low, W., Harries, M., Ye, H., Du, M. Q., Boshoff, C. & Collins, M. (2001). Internal ribosome entry site regulates translation of Kaposi's sarcoma-associated herpesvirus FLICE inhibitory protein. *J Virol* **75**, 2938-2945.

Lu, P. D., Harding, H. P. & Ron, D. (2004). Translation reinitiation at alternative open reading frames regulates gene expression in an integrated stress response. *J Cell Biol* **167**, 27-33.

Lucchesi, W., Brady, G., Dittrich-Breiholz, O., Kracht, M., Russ, R. & Farrell, P. J. (2008). Differential gene regulation by Epstein-Barr virus type 1 and type 2 EBNA2. *J Virol* **82**, 7456-7466.

Lukac, D. M., Renne, R., Kirshner, J. R. & Ganem, D. (1998). Reactivation of Kaposi's sarcoma-associated herpesvirus infection from latency by expression of the ORF 50 transactivator, a homolog of the EBV R protein. *Virology* **252**, 304-312.

Maestre, L., Tooze, R., Canamero, M., Montes-Moreno, S., Ramos, R., Doody, G., Boll, M., Barrans, S., Baena, S., Piris, M. A. & Roncador, G. (2009). Expression pattern of XBP1(S) in human B-cell lymphomas. *Haematologica* **94**, 419-422.

Malatack, J. F., Gartner, J. C., Jr., Urbach, A. H. & Zitelli, B. J. (1991). Orthotopic liver transplantation, Epstein-Barr virus, cyclosporine, and lymphoproliferative disease: a growing concern. *J Pediatr* **118**, 667-675.

Malin, S., Linderson, Y., Almqvist, J., Ernberg, I., Tallone, T. & Pettersson, S. (2005). DNA-dependent conversion of Oct-1 and Oct-2 into transcriptional repressors by Groucho/TLE. *Nucleic Acids Res* **33**, 4618-4625.

Manet, E., Gruffat, H., Trescol-Biemont, M. C., Moreno, N., Chambard, P., Giot, J. F. & Sergeant, A. (1989). Epstein-Barr virus bicistronic mRNAs generated

by facultative splicing code for two transcriptional trans-activators. *EMBO J* **8**, 1819-1826.

Manet, E., Rigolet, A., Gruffat, H., Giot, J. F. & Sergeant, A. (1991). Domains of the Epstein-Barr virus (EBV) transcription factor R required for dimerization, DNA binding and activation. *Nucleic Acids Res* **19**, 2661-2667.

Manolov, G. & Manolova, Y. (1972). Marker band in one chromosome 14 from Burkitt lymphomas. *Nature* **237**, 33-34.

Mansky, L. M. & Temin, H. M. (1995). Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol* **69**, 5087-5094.

Marciniak, S. J., Yun, C. Y., Oyadomari, S., Novoa, I., Zhang, Y., Jungreis, R., Nagata, K., Harding, H. P. & Ron, D. (2004). CHOP induces death by promoting protein synthesis and oxidation in the stressed endoplasmic reticulum. *Genes Dev* **18**, 3066-3077.

Martin, D. & Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. *Bioinformatics* **16**, 562-563.

Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D. & Lefeuvre, P. (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26**, 2462-2463.

Martin, D. P., Posada, D., Crandall, K. A. & Williamson, C. (2005). A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res Hum Retroviruses* **21**, 98-102.

Martin, F., Oliver, A. M. & Kearney, J. F. (2001). Marginal zone and B1 B cells unite in the early response against T-independent blood-borne particulate antigens. *Immunity* **14**, 617-629.

Martini, M., Capello, D., Serraino, D., Navarra, A., Pierconti, F., Cenci, T., Gaidano, G. & Larocca, L. M. (2007). Characterization of variants in the promoter of EBV gene BZLF1 in normal donors, HIV-positive patients and in AIDS-related lymphomas. *J Infect* **54**, 298-306.

Maruo, S., Wu, Y., Ishikawa, S., Kanda, T., Iwakiri, D. & Takada, K. (2006). Epstein-Barr virus nuclear protein EBNA3C is required for cell cycle progression and growth maintenance of lymphoblastoid cells. *Proc Natl Acad Sci U S A* **103**, 19500-19505.

Masters, J. R. (2002). HeLa cells 50 years on: the good, the bad and the ugly. *Nat Rev Cancer* **2**, 315-319.

Matolcsy, A., Nador, R. G., Cesarman, E. & Knowles, D. M. (1998). Immunoglobulin VH gene mutational analysis suggests that primary effusion lymphomas derive from different stages of B cell maturation. *Am J Pathol* **153**, 1609-1614.

Matthews, S. A., Rozengurt, E. & Cantrell, D. (2000). Protein kinase D. A selective target for antigen receptors and a downstream target for protein kinase C in lymphocytes. *J Exp Med* **191**, 2075-2082.

Mbulaiteye, S. M., Biggar, R. J., Bhatia, K., Linet, M. S. & Devesa, S. S. (2009). Sporadic childhood Burkitt lymphoma incidence in the United States during 1992-2005. *Pediatr Blood Cancer* **53**, 366-370.

McDonald, C., Karstegl, C. E., Kellam, P. & Farrell, P. J. (2010). Regulation of the Epstein-Barr virus Zp promoter in B lymphocytes during reactivation from latency. *J Gen Virol* **91**, 622-629.

McGeoch, D. J. & Davison, A. J. (1999). The descent of human herpesvirus 8. *Semin Cancer Biol* **9**, 201-209.

McHeyzer-Williams, L. J., Cool, M. & McHeyzer-Williams, M. G. (2000). Antigen-specific B cell memory: expression and replenishment of a novel b220(-) memory b cell compartment. *J Exp Med* **191**, 1149-1166.

McKean, D., Huppi, K., Bell, M., Staudt, L., Gerhard, W. & Weigert, M. (1984). Generation of antibody diversity in the immune response of BALB/c mice to influenza virus hemagglutinin. *Proc Natl Acad Sci U S A* **81**, 3180-3184.

McNiece, I. K., Langley, K. E. & Zsebo, K. M. (1991). Recombinant human stem cell factor synergises with GM-CSF, G-CSF, IL-3 and epo to stimulate human progenitor cells of the myeloid and erythroid lineages. *Exp Hematol* **19**, 226-231.

Mehr, R., Shannon, M. & Litwin, S. (1999). Models for antigen receptor gene rearrangement. I. Biased receptor editing in B cells: implications for allelic exclusion. *J Immunol* **163**, 1793-1798.

Melchers, F., ten Boekel, E., Seidl, T., Kong, X. C., Yamagami, T., Onishi, K., Shimizu, T., Rolink, A. G. & Andersson, J. (2000). Repertoire selection by pre-B-cell receptors and B-cell receptors, and genetic control of B-cell development from immature to mature B cells. *Immunol Rev* **175**, 33-46.

Meng, Y. X., Spira, T. J., Bhat, G. J., Birch, C. J., Druce, J. D., Edlin, B. R., Edwards, R., Gunthel, C., Newton, R., Stamey, F. R., Wood, C. & Pellett, P. E. (1999). Individuals from North America, Australasia, and Africa are infected with four different genotypes of human herpesvirus 8. *Virology* **261**, 106-119.

Mettenleiter, T. C. (2004). Budding events in herpesvirus morphogenesis. *Virus Res* **106**, 167-180.

Midgley, R. S., Bell, A. I., McGeoch, D. J. & Rickinson, A. B. (2003). Latent gene sequencing reveals familial relationships among Chinese Epstein-Barr virus strains and evidence for positive selection of A11 epitope changes. *J Virol* **77**, 11517-11530.

Midgley, R. S., Blake, N. W., Yao, Q. Y., Croom-Carter, D., Cheung, S. T., Leung, S. F., Chan, A. T., Johnson, P. J., Huang, D., Rickinson, A. B. & Lee, S. P. (2000). Novel intertypic recombinants of epstein-barr virus in the chinese population. *J Virol* **74**, 1544-1548.

Miller, G., Heston, L., Grogan, E., Gradoville, L., Rigsby, M., Sun, R., Shedd, D., Kushnaryov, V. M., Grossberg, S. & Chang, Y. (1997). Selective switch between latency and lytic replication of Kaposi's sarcoma herpesvirus and Epstein-Barr virus in dually infected body cavity lymphoma cells. *J Virol* **71**, 314-324.

Miller, G., Shope, T., Lisco, H., Stitt, D. & Lipman, M. (1972). Epstein-Barr virus: transformation, cytopathic changes, and viral antigens in squirrel monkey and marmoset leukocytes. *Proc Natl Acad Sci U S A* **69**, 383-387.

Miller, I. G., Jr. & El-Guindy, A. (2002). Regulation of Epstein-Barr virus lytic cycle activation in malignant and nonmalignant disease. *J Natl Cancer Inst* **94**, 1733-1735.

Miller, N. & Hutt-Fletcher, L. M. (1988). A monoclonal antibody to glycoprotein gp85 inhibits fusion but not attachment of Epstein-Barr virus. *J Virol* **62**, 2366-2372.

Miller, N. & Hutt-Fletcher, L. M. (1992). Epstein-Barr virus enters B cells and epithelial cells by different routes. *J Virol* **66**, 3409-3414.

Mitchell, G. F., Chan, E. L., Noble, M. S., Weissman, I. L., Mishell, R. I. & Herzenberg, L. A. (1972). Immunological memory in mice. 3. Memory to heterologous erythrocytes in both T cell and B cell populations and requirement for T cells in expression of B cell memory. Evidence using

immunoglobulin allotype and mouse alloantigen theta markers with congenic mice. *J Exp Med* **135**, 165-184.

Moore, P. S., Gao, S. J., Dominguez, G., Cesarman, E., Lungu, O., Knowles, D. M., Garber, R., Pellett, P. E., McGeoch, D. J. & Chang, Y. (1996). Primary characterization of a herpesvirus agent associated with Kaposi's sarcomae. *J Virol* **70**, 549-558.

Mori, S., Itoh, T., Tokunaga, M. & Eizuru, Y. (1999). Deletions and single-base mutations within the carboxy-terminal region of the latent membrane protein 1 oncogene in Epstein-Barr virus-related gastric cancers of southern Japan. *J Med Virol* **57**, 152-158.

Mueller, N. (1999). Overview of the epidemiology of malignancy in immune deficiency. *J Acquir Immune Defic Syndr* **21 Suppl 1**, S5-10.

Muralidhar, S., Veytsmann, G., Chandran, B., Ablashi, D., Doniger, J. & Rosenthal, L. J. (2000). Characterization of the human herpesvirus 8 (Kaposi's sarcoma-associated herpesvirus) oncogene, kaposin (ORF K12). *J Clin Virol* **16**, 203-213.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y. & Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553-563.

Murata, K., Ishii, N., Takano, H., Miura, S., Ndhlovu, L. C., Nose, M., Noda, T. & Sugamura, K. (2000). Impairment of antigen-presenting cell function in mice lacking expression of OX40 ligand. *J Exp Med* **191**, 365-374.

Murata, T., Kondo, Y., Sugimoto, A., Kawashima, D., Saito, S., Isomura, H., Kanda, T. & Tsurumi, T. (2012). Epigenetic Histone Modification of Epstein-Barr virus BZLF1 Promoter during Latency and Reactivation in Raji Cells. *J Virol*.

Nador, R. G., Cesarman, E., Chadburn, A., Dawson, D. B., Ansari, M. Q., Sald, J. & Knowles, D. M. (1996). Primary effusion lymphoma: a distinct clinicopathologic entity associated with the Kaposi's sarcoma-associated herpes virus. *Blood* **88**, 645-656.

Nagamine, M., Takahara, M., Kishibe, K., Nagato, T., Ishii, H., Bandoh, N., Ogino, T. & Harabuchi, Y. (2007). Sequence variations of Epstein-Barr virus LMP1 gene in nasal NK/T-cell lymphoma. *Virus Genes* **34**, 47-54.

Nagasawa, T., Hirota, S., Tachibana, K., Takakura, N., Nishikawa, S., Kitamura, Y., Yoshida, N., Kikutani, H. & Kishimoto, T. (1996). Defects of B-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the CXC chemokine PBSF/SDF-1. *Nature* **382**, 635-638.

Nalesnik, M. A. (1996). Posttransplantation lymphoproliferative disorders (PTLD): current perspectives. *Semin Thorac Cardiovasc Surg* **8**, 139-148.

Nalesnik, M. A. (1998). Clinical and pathological features of post-transplant lymphoproliferative disorders (PTLD). *Springer Semin Immunopathol* **20**, 325-342.

Nazerian, K. (1974). DNA configuration in the core of Marek's disease virus. *J Virol* **13**, 1148-1150.

Neipel, F., Albrecht, J. C. & Fleckenstein, B. (1998). Human herpesvirus 8--the first human Rhadinovirus. *J Natl Cancer Inst Monogr*, 73-77.

Nemazee, D. A. & Burki, K. (1989). Clonal deletion of B lymphocytes in a transgenic mouse bearing anti-MHC class I antibody genes. *Nature* **337**, 562-566.

Nemerow, G. R. & Cooper, N. R. (1984). Early events in the infection of human B lymphocytes by Epstein-Barr virus: the internalization process. *Virology* **132**, 186-198.

Newell, K. A., Alonso, E. M., Whitington, P. F., Bruce, D. S., Millis, J. M., Piper, J. B., Woodle, E. S., Kelly, S. M., Koeppen, H., Hart, J., Rubin, C. M. & Thistlethwaite, J. R., Jr. (1996). Posttransplant lymphoproliferative disease in pediatric liver transplantation. Interplay between primary Epstein-Barr virus infection and immunosuppression. *Transplantation* **62**, 370-375.

Newman, J. R. & Keating, A. E. (2003). Comprehensive identification of human bZIP interactions with coiled-coil arrays. *Science* **300**, 2097-2101.

Nishitoh, H., Matsuzawa, A., Tobiume, K., Saegusa, K., Takeda, K., Inoue, K., Hori, S., Kakizuka, A. & Ichijo, H. (2002). ASK1 is essential for endoplasmic reticulum stress-induced neuronal cell death triggered by expanded polyglutamine repeats. *Genes Dev* **16**, 1345-1355.

Nitsche, F., Bell, A. & Rickinson, A. (1997). Epstein-Barr virus leader protein enhances EBNA-2-mediated transactivation of latent membrane protein 1 expression: a role for the W1W2 repeat domain. *J Virol* **71**, 6619-6628.

Nonkwelo, C., Skinner, J., Bell, A., Rickinson, A. & Sample, J. (1996). Transcription start sites downstream of the Epstein-Barr virus (EBV) Fp promoter in early-passage Burkitt lymphoma cells define a fourth promoter for expression of the EBV EBNA-1 protein. *J Virol* **70**, 623-627.

Norseen, J., Thomae, A., Sridharan, V., Aiyar, A., Schepers, A. & Lieberman, P. M. (2008). RNA-dependent recruitment of the origin recognition complex. *EMBO J* **27**, 3024-3035.

Norvell, A., Mandik, L. & Monroe, J. G. (1995). Engagement of the antigen-receptor on immature murine B lymphocytes results in death by apoptosis. *J Immunol* **154**, 4404-4413.

Nutt, S. L., Eberhard, D., Horcher, M., Rolink, A. G. & Busslinger, M. (2001). Pax5 determines the identity of B cells from the beginning to the end of B-lymphopoiesis. *Int Rev Immunol* **20**, 65-82.

Nutt, S. L., Rolink, A. G. & Busslinger, M. (1999). The molecular basis of B-cell lineage commitment. *Cold Spring Harb Symp Quant Biol* **64**, 51-59.

Oettinger, M. A., Schatz, D. G., Gorka, C. & Baltimore, D. (1990). RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* **248**, 1517-1523.

Ogata, M., Hino, S., Saito, A., Morikawa, K., Kondo, S., Kanemoto, S., Murakami, T., Taniguchi, M., Tanii, I., Yoshinaga, K., Shiosaka, S., Hammarback, J. A., Urano, F. & Imaizumi, K. (2006). Autophagy is activated for cell survival after endoplasmic reticulum stress. *Mol Cell Biol* **26**, 9220-9231.

Okada, T., Yoshida, H., Akazawa, R., Negishi, M. & Mori, K. (2002). Distinct roles of activating transcription factor 6 (ATF6) and double-stranded RNA-activated protein kinase-like endoplasmic reticulum kinase (PERK) in transcription during the mammalian unfolded protein response. *Biochem J* **366**, 585-594.

Okkenhaug, K. & Vanhaesebroeck, B. (2003). PI3K in lymphocyte development, differentiation and activation. *Nat Rev Immunol* **3**, 317-330.

Ollila, J. & Vihinen, M. (2005). B cells. *Int J Biochem Cell Biol* **37**, 518-523.

Packham, G., Brimmell, M., Cook, D., Sinclair, A. J. & Farrell, P. J. (1993). Strain variation in Epstein-Barr virus immediate early genes. *Virology* **192**, 541-550.

Palm, W. & de Lange, T. (2008). How shelterin protects mammalian telomeres. *Annu Rev Genet* **42**, 301-334.

Panagopoulos, D., Victoratos, P., Alexiou, M., Kollias, G. & Mosialos, G. (2004). Comparative analysis of signal transduction by CD40 and the Epstein-Barr virus oncoprotein LMP1 in vivo. *J Virol* **78**, 13253-13261.

Parekh, S., Polo, J. M., Shaknovich, R., Juszczynski, P., Lev, P., Ranuncolo, S. M., Yin, Y., Klein, U., Cattoretti, G., Dalla Favera, R., Shipp, M. A. & Melnick, A. (2007). BCL6 programs lymphoma cells for survival and differentiation through distinct biochemical mechanisms. *Blood* **110**, 2067-2074.

Parker, G. A., Crook, T., Bain, M., Sara, E. A., Farrell, P. J. & Allday, M. J. (1996). Epstein-Barr virus nuclear antigen (EBNA)3C is an immortalizing oncoprotein with similar properties to adenovirus E1A and papillomavirus E7. *Oncogene* **13**, 2541-2549.

Paulson, E. J. & Speck, S. H. (1999). Differential methylation of Epstein-Barr virus latency promoters facilitates viral persistence in healthy seropositive individuals. *J Virol* **73**, 9959-9968.

Peng, C. W., Xue, Y., Zhao, B., Johannsen, E., Kieff, E. & Harada, S. (2004). Direct interactions between Epstein-Barr virus leader protein LP and the EBNA2 acidic domain underlie coordinate transcriptional regulation. *Proc Natl Acad Sci U S A* **101**, 1033-1038.

Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M. & Ho, D. D. (1996). HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* **271**, 1582-1586.

Pfeffer, S., Sewer, A., Lagos-Quintana, M., Sheridan, R., Sander, C., Grasser, F. A., van Dyk, L. F., Ho, C. K., Shuman, S., Chien, M., Russo, J. J., Ju, J., Randall, G., Lindenbach, B. D., Rice, C. M., Simon, V., Ho, D. D., Zavolan, M. & Tuschl, T. (2005). Identification of microRNAs of the herpesvirus family. *Nat Methods* **2**, 269-276.

Philip, T. (1985). Burkitt's lymphoma in Europe. *IARC Sci Publ*, 107-118.

Pickhardt, P. J., Siegel, M. J., Anderson, D. C., Hayashi, R. & DeBaun, M. R. (1998). Chest radiography as a predictor of outcome in posttransplantation lymphoproliferative disorder in lung allograft recipients. *AJR Am J Roentgenol* **171**, 375-382.

Pickhardt, P. J., Siegel, M. J., Hayashi, R. J. & Kelly, M. (2000). Posttransplantation lymphoproliferative disorder in children: clinical, histopathologic, and imaging features. *Radiology* **217**, 16-25.

Pillai, S. & Cariappa, A. (2009). The follicular versus marginal zone B lymphocyte cell fate decision. *Nat Rev Immunol* **9**, 767-777.

Platt, G. M., Cannell, E., Cuomo, M. E., Singh, S. & Mittnacht, S. (2000). Detection of the human herpesvirus 8-encoded cyclin protein in primary effusion lymphoma-derived cell lines. *Virology* **272**, 257-266.

Poole, L. J., Zong, J. C., Ciufo, D. M., Alcendor, D. J., Cannon, J. S., Ambinder, R., Orenstein, J. M., Reitz, M. S. & Hayward, G. S. (1999). Comparison of genetic variability at multiple loci across the genomes of the major subtypes of Kaposi's sarcoma-associated herpesvirus reveals evidence for recombination and for two distinct types of open reading frame K15 alleles at the right-hand end. *J Virol* **73**, 6646-6660.

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Mol Biol Evol* **25**, 1253-1256.

Posada, D. & Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci U S A* **98**, 13757-13762.

Prince, S., Keating, S., Fielding, C., Brennan, P., Floettmann, E. & Rowe, M. (2003). Latent membrane protein 1 inhibits Epstein-Barr virus lytic cycle induction and progress *via* different mechanisms. *J Virol* **77**, 5000-5007.

Puglielli, M. T., Woisetschlaeger, M. & Speck, S. H. (1996). oriP is essential for EBNA gene promoter activity in Epstein-Barr virus-immortalized lymphoblastoid cell lines. *J Virol* **70**, 5758-5768.

Pule, M. A., Savoldo, B., Myers, G. D., Rossig, C., Russell, H. V., Dotti, G., Huls, M. H., Liu, E., Gee, A. P., Mei, Z., Yvon, E., Weiss, H. L., Liu, H., Rooney, C. M., Heslop, H. E. & Brenner, M. K. (2008). Virus-specific T cells engineered to coexpress tumor-specific receptors: persistence and antitumor activity in individuals with neuroblastoma. *Nat Med* **14**, 1264-1270.

Pulvertaft, J. V. (1964). Cytology of Burkitt's Tumour (African Lymphoma). *Lancet* **1**, 238-240.

Raab-Traub, N. & Flynn, K. (1986). The structure of the termini of the Epstein-Barr virus as a marker of clonal cellular proliferation. *Cell* **47**, 883-889.

Rada, C., Williams, G. T., Nilsen, H., Barnes, D. E., Lindahl, T. & Neuberger, M. S. (2002). Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr Biol* **12**, 1748-1755.

Radkov, S. A., Bain, M., Farrell, P. J., West, M., Rowe, M. & Allday, M. J. (1997). Epstein-Barr virus EBNA3C represses Cp, the major promoter for EBNA expression, but has no effect on the promoter of the cell gene CD21. *J Virol* **71**, 8552-8562.

Radkov, S. A., Touitou, R., Brehm, A., Rowe, M., West, M., Kouzarides, T. & Allday, M. J. (1999). Epstein-Barr virus nuclear antigen 3C interacts with histone deacetylase to repress transcription. *J Virol* **73**, 5688-5697.

Raggo, C., Ruhl, R., McAllister, S., Koon, H., Dezube, B. J., Fruh, K. & Moses, A. V. (2005). Novel cellular genes essential for transformation of endothelial cells by Kaposi's sarcoma-associated herpesvirus. *Cancer Res* **65**, 5084-5095.

Ragoczy, T., Heston, L. & Miller, G. (1998). The Epstein-Barr virus Rta protein activates lytic cycle genes and can disrupt latency in B lymphocytes. *J Virol* **72**, 7978-7984.

Rainbow, L., Platt, G. M., Simpson, G. R., Sarid, R., Gao, S. J., Stoiber, H., Herrington, C. S., Moore, P. S. & Schulz, T. F. (1997). The 222- to 234-kilodalton latent nuclear protein (LNA) of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) is encoded by orf73 and is a component of the latency-associated nuclear antigen. *J Virol* **71**, 5915-5921.

Rawlins, D. R., Milman, G., Hayward, S. D. & Hayward, G. S. (1985). Sequence-specific DNA binding of the Epstein-Barr virus nuclear antigen (EBNA-1) to clustered sites in the plasmid maintenance region. *Cell* **42**, 859-868.

Razzouk, B. I., Srinivas, S., Sample, C. E., Singh, V. & Sixbey, J. W. (1996). Epstein-Barr Virus DNA recombination and loss in sporadic Burkitt's lymphoma. *J Infect Dis* **173**, 529-535.

Rechsteiner, M. P., Bernasconi, M., Berger, C. & Nadal, D. (2008). Role of latent membrane protein 2 isoforms in Epstein-Barr virus latency. *Trends Microbiol* **16**, 520-527.

Reed, J. A., Nador, R. G., Spaulding, D., Tani, Y., Cesarman, E. & Knowles, D. M. (1998). Demonstration of Kaposi's sarcoma-associated herpes virus cyclin

D homolog in cutaneous Kaposi's sarcoma by colorimetric in situ hybridization using a catalyzed signal amplification system. *Blood* **91**, 3825-3832.

Reimold, A. M., Iwakoshi, N. N., Manis, J., Vallabhajosyula, P., Szomolanyi-Tsuda, E., Gravallese, E. M., Friend, D., Grusby, M. J., Alt, F. & Glimcher, L. H. (2001). Plasma cell differentiation requires the transcription factor XBP-1. *Nature* **412**, 300-307.

Reimold, A. M., Ponath, P. D., Li, Y. S., Hardy, R. R., David, C. S., Strominger, J. L. & Glimcher, L. H. (1996). Transcription factor B cell lineage-specific activator protein regulates the gene for human X-box binding protein 1. *J Exp Med* **183**, 393-401.

Reisman, D. & Sugden, B. (1986). trans activation of an Epstein-Barr viral transcriptional enhancer by the Epstein-Barr viral nuclear antigen 1. *Mol Cell Biol* **6**, 3838-3846.

Reisman, D., Yates, J. & Sugden, B. (1985). A putative origin of replication of plasmids derived from Epstein-Barr virus is composed of two cis-acting components. *Mol Cell Biol* **5**, 1822-1832.

Renne, R., Lagunoff, M., Zhong, W. & Ganem, D. (1996a). The size and conformation of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) DNA in infected cells and virions. *J Virol* **70**, 8151-8154.

Renne, R., Zhong, W., Herndier, B., McGrath, M., Abbey, N., Kedes, D. & Ganem, D. (1996b). Lytic growth of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in culture. *Nat Med* **2**, 342-346.

Renzette, N., Bhattacharjee, B., Jensen, J. D., Gibson, L. & Kowalik, T. F. (2011). Extensive genome-wide variability of human cytomegalovirus in congenitally infected infants. *PLoS Pathog* **7**, e1001344.

Ressing, M. E., van Leeuwen, D., Verreck, F. A., Keating, S., Gomez, R., Franken, K. L., Ottenhoff, T. H., Spriggs, M., Schumacher, T. N., Hutt-Fletcher, L. M., Rowe, M. & Wiertz, E. J. (2005). Epstein-Barr virus gp42 is posttranslationally modified to produce soluble gp42 that mediates HLA class II immune evasion. *J Virol* **79**, 841-852.

Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Labelouse, R., Gennery, A., Tezcan, I., Ersoy, F., Kayserili, H., Ugazio, A. G., Brousse, N., Muramatsu, M., Notarangelo, L. D., Kinoshita, K., Honjo, T., Fischer, A. & Durandy, A. (2000). Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell* **102**, 565-575.

Rey, O., Papazyan, R., Waldron, R. T., Young, S. H., Lippincott-Schwartz, J., Jacamo, R. & Rozengurt, E. (2006). The nuclear import of protein kinase D3 requires its catalytic activity. *J Biol Chem* **281**, 5149-5157.

Rickinson, A. & Kieff, E. (1996). Field's Virology, third Edition pp. 2343-2380. Edited by P. M. Howley: Lippincott-Raven: Philadelphia.

Rickinson, A. & Kieff, E. (2001). *Fields Virology*: Lippincott Williams & Wilkins Publishers: Pjiladelphia.

Rickinson, A. & Kieff, E. (2006). Epstein-Barr Virus. In *Fields Virology*, 4th Edtion edn. Edited by D. M. Knipe, P. M. Howley, D. E. Griffin, R. A. Lamb, M. A. Martin, B. Roizman & S. E. Straus: Lippincott Williams & Wilkins, Philadelphia, PA.

Rickinson, A. B., Young, L. S. & Rowe, M. (1987). Influence of the Epstein-Barr virus nuclear antigen EBNA 2 on the growth phenotype of virus-transformed B cells. *J Virol* **61**, 1310-1317.

Rivas, C., Thlick, A. E., Parravicini, C., Moore, P. S. & Chang, Y. (2001). Kaposi's sarcoma-associated herpesvirus LANA2 is a B-cell-specific latent viral protein that inhibits p53. *J Virol* **75**, 429-438.

Robertson, D. L., Hahn, B. H. & Sharp, P. M. (1995a). Recombination in AIDS viruses. *J Mol Evol* **40**, 249-259.

Robertson, E. S., Grossman, S., Johannsen, E., Miller, C., Lin, J., Tomkinson, B. & Kieff, E. (1995b). Epstein-Barr virus nuclear protein 3C modulates transcription through interaction with the sequence-specific DNA-binding protein J kappa. *J Virol* **69**, 3108-3116.

Robertson, K. D. & Ambinder, R. F. (1997). Methylation of the Epstein-Barr virus genome in normal lymphocytes. *Blood* **90**, 4480-4484.

Robinson, A. R., Kwek, S. S., Hagemeier, S. R., Wille, C. K. & Kenney, S. C. (2011). Cellular transcription factor Oct-1 interacts with the Epstein-Barr virus BRLF1 protein to promote disruption of viral latency. *J Virol* **85**, 8940-8953.

Robinson, A. R., Kwek, S. S. & Kenney, S. C. (2012). The B-Cell Specific Transcription Factor, Oct-2, Promotes Epstein-Barr Virus Latency by Inhibiting the Viral Immediate-Early Protein, BZLF1. *PLoS Pathog* **8**, e1002516.

Rochford, R., Cannon, M. J. & Moormann, A. M. (2005). Endemic Burkitt's lymphoma: a polymicrobial disease? *Nat Rev Microbiol* **3**, 182-187.

Roizman, B. & Whitley, R. J. (2001). The nine ages of herpes simplex virus. *Herpes* **8**, 23-27.

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M. & Nyren, P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem* **242**, 84-89.

Rooney, C., Howe, J. G., Speck, S. H. & Miller, G. (1989a). Influence of Burkitt's lymphoma and primary B cells on latent gene expression by the nonimmortalizing P3J-HR-1 strain of Epstein-Barr virus. *J Virol* **63**, 1531-1539.

Rooney, C. M., Brimmell, M., Buschle, M., Allan, G., Farrell, P. J. & Kolman, J. L. (1992). Host cell and EBNA-2 regulation of Epstein-Barr virus latent-cycle promoter activity in B lymphocytes. *J Virol* **66**, 496-504.

Rooney, C. M., Rowe, D. T., Ragot, T. & Farrell, P. J. (1989b). The spliced BZLF1 gene of Epstein-Barr virus (EBV) transactivates an early EBV promoter and induces the virus productive cycle. *J Virol* **63**, 3109-3116.

Rovedo, M. & Longnecker, R. (2007). Epstein-barr virus latent membrane protein 2B (LMP2B) modulates LMP2A activity. *J Virol* **81**, 84-94.

Rowe, M., Kelly, G. L., Bell, A. I. & Rickinson, A. B. (2009). Burkitt's lymphoma: the Rosetta Stone deciphering Epstein-Barr virus biology. *Semin Cancer Biol* **19**, 377-388.

Rowe, M., Rowe, D. T., Gregory, C. D., Young, L. S., Farrell, P. J., Rupani, H. & Rickinson, A. B. (1987). Differences in B cell growth phenotype reflect novel patterns of Epstein-Barr virus latent gene expression in Burkitt's lymphoma cells. *EMBO J* **6**, 2743-2751.

Russell, D. M., Dembic, Z., Morahan, G., Miller, J. F., Burki, K. & Nemazee, D. (1991). Peripheral deletion of self-reactive B cells. *Nature* **354**, 308-311.

Russo, J. J., Bohenzky, R. A., Chien, M. C., Chen, J., Yan, M., Maddalena, D., Parry, J. P., Peruzzi, D., Edelman, I. S., Chang, Y. & Moore, P. S. (1996). Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8). *Proc Natl Acad Sci U S A* **93**, 14862-14867.

Rutkowski, D. T., Arnold, S. M., Miller, C. N., Wu, J., Li, J., Gunnison, K. M., Mori, K., Sadighi Akha, A. A., Raden, D. & Kaufman, R. J. (2006). Adaptation to ER stress is mediated by differential stabilities of pro-survival and pro-apoptotic mRNAs and proteins. *PLoS Biol* **4**, e374.

Sacaze, C., Henry, S., Icart, J. & Mariame, B. (2001). Tissue specific distribution of Epstein-Barr virus (EBV) BZLF1 gene variants in nasopharyngeal carcinoma (NPC) bearing patients. *Virus Res* **81**, 133-142.

Saechan, V., Mori, A., Mitarnun, W., Settheetham-Ishida, W. & Ishida, T. (2006). Analysis of LMP1 variants of EBV in Southern Thailand: evidence for strain-associated T-cell tropism and pathogenicity. *J Clin Virol* **36**, 119-125.

Sairenji, T., Daibata, M., Sorli, C. H., Qvistback, H., Humphreys, R. E., Ludvigsson, J., Palmer, J., Landin-Olsson, M., Sundkvist, G., Michelsen, B. & et al. (1991). Relating homology between the Epstein-Barr virus BOLF1 molecule and HLA-DQw8 beta chain to recent onset type 1 (insulin-dependent) diabetes mellitus. *Diabetologia* **34**, 33-39.

Samols, M. A., Hu, J., Skalsky, R. L. & Renne, R. (2005). Cloning and identification of a microRNA cluster within the latency-associated region of Kaposi's sarcoma-associated herpesvirus. *J Virol* **79**, 9301-9305.

Sample, J., Hummel, M., Braun, D., Birkenbach, M. & Kieff, E. (1986). Nucleotide sequences of mRNAs encoding Epstein-Barr virus nuclear proteins: a probable transcriptional initiation site. *Proc Natl Acad Sci U S A* **83**, 5096-5100.

Sample, J., Young, L., Martin, B., Chatman, T., Kieff, E. & Rickinson, A. (1990). Epstein-Barr virus types 1 and 2 differ in their EBNA-3A, EBNA-3B, and EBNA-3C genes. *J Virol* **64**, 4084-4092.

Sandvej, K., Zhou, X. G. & Hamilton-Dutoit, S. (2000). EBNA-1 sequence variation in Danish and Chinese EBV-associated tumours: evidence for geographical polymorphism but not for tumour-specific subtype restriction. *J Pathol* **191**, 127-131.

Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, C. A., Hutchison, C. A., Slocombe, P. M. & Smith, M. (1977a). Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* **265**, 687-695.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977b). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463-5467.

Sarisky, R. T., Gao, Z., Lieberman, P. M., Fixman, E. D., Hayward, G. S. & Hayward, S. D. (1996). A replication function associated with the activation domain of the Epstein-Barr virus Zta transactivator. *J Virol* **70**, 8340-8347.

Sarrias, M. R., Franchini, S., Canziani, G., Argyropoulos, E., Moore, W. T., Sahu, A. & Lambris, J. D. (2001). Kinetic analysis of the interactions of complement receptor 2 (CR2, CD21) with its ligands C3d, iC3b, and the EBV glycoprotein gp350/220. *J Immunol* **167**, 1490-1499.

Savard, M., Belanger, C., Tardif, M., Gourde, P., Flamand, L. & Gosselin, J. (2000). Infection of primary human monocytes by Epstein-Barr virus. *J Virol* **74**, 2612-2619.

Schaefer, B. C., Strominger, J. L. & Speck, S. H. (1995). Redefining the Epstein-Barr virus-encoded nuclear antigen EBNA-1 gene promoter and transcription initiation site in group I Burkitt lymphoma cell lines. *Proc Natl Acad Sci U S A* **92**, 10565-10569.

Schelcher, C., Valencia, S., Delecluse, H. J., Hicks, M. & Sinclair, A. J. (2005). Mutation of a single amino acid residue in the basic region of the Epstein-Barr virus (EBV) lytic cycle switch protein Zta (BZLF1) prevents reactivation of EBV from latency. *J Virol* **79**, 13822-13828.

Schepers, A., Pich, D. & Hammerschmidt, W. (1993a). A transcription factor with homology to the AP-1 family links RNA transcription and DNA replication in the lytic cycle of Epstein-Barr virus. *EMBO J* **12**, 3921-3929.

Schepers, A., Pich, D. & Hammerschmidt, W. (1996). Activation of oriLyt, the lytic origin of DNA replication of Epstein-Barr virus, by BZLF1. *Virology* **220**, 367-376.

Schepers, A., Pich, D., Mankertz, J. & Hammerschmidt, W. (1993b). cis-acting elements in the lytic origin of DNA replication of Epstein-Barr virus. *J Virol* **67**, 4237-4245.

Schepers, A., Ritzi, M., Bousset, K., Kremmer, E., Yates, J. L., Harwood, J., Diffley, J. F. & Hammerschmidt, W. (2001). Human origin recognition complex binds to the region of the latent origin of DNA replication of Epstein-Barr virus. *EMBO J* **20**, 4588-4602.

Schlager, S., Speck, S. H. & Woisetschlager, M. (1996). Transcription of the Epstein-Barr virus nuclear antigen 1 (EBNA1) gene occurs before induction of the BCR2 (Cp) EBNA gene promoter during the initial stages of infection in B cells. *J Virol* **70**, 3561-3570.

Schliephake, D. E. & Schimpl, A. (1996). Blimp-1 overcomes the block in IgM secretion in lipopolysaccharide/anti-mu F(ab')2-co-stimulated B lymphocytes. *Eur J Immunol* **26**, 268-271.

Schmaus, S., Wolf, H. & Schwarzmann, F. (2004). The reading frame BPLF1 of Epstein-Barr virus: a homologue of herpes simplex virus protein VP16. *Virus Genes* **29**, 267-277.

Schmidlin, H., Diehl, S. A. & Blom, B. (2009). New insights into the regulation of human B-cell differentiation. *Trends Immunol* **30**, 277-285.

Schroder, M. & Kaufman, R. J. (2005). ER stress and the unfolded protein response. *Mutat Res* **569**, 29-63.

Sears, J., Ujihara, M., Wong, S., Ott, C., Middeldorp, J. & Aiyar, A. (2004). The amino terminus of Epstein-Barr Virus (EBV) nuclear antigen 1 contains AT hooks that facilitate the replication and partitioning of latent EBV genomes by tethering them to cellular chromosomes. *J Virol* **78**, 11487-11505.

Seigneurin, J. M., Vuillaume, M., Lenoir, G. & De-The, G. (1977). Replication of Epstein-Barr virus: ultrastructural and immunofluorescent studies of P3HR1-superinfected Raji cells. *J Virol* **24**, 836-845.

Shaffer, A. L., Lin, K. I., Kuo, T. C., Yu, X., Hurt, E. M., Rosenwald, A., Giltnane, J. M., Yang, L., Zhao, H., Calame, K. & Staudt, L. M. (2002). Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program. *Immunity* **17**, 51-62.

Shaffer, A. L., Shapiro-Shelef, M., Iwakoshi, N. N., Lee, A. H., Qian, S. B., Zhao, H., Yu, X., Yang, L., Tan, B. K., Rosenwald, A., Hurt, E. M., Petroulakis, E., Sonenberg, N., Yewdell, J. W., Calame, K., Glimcher, L. H. & Staudt, L. M. (2004). XBP1, downstream of Blimp-1, expands the secretory apparatus and other organelles, and increases protein synthesis in plasma cell differentiation. *Immunity* **21**, 81-93.

Shaffer, A. L., Yu, X., He, Y., Boldrick, J., Chan, E. P. & Staudt, L. M. (2000). BCL-6 represses genes that function in lymphocyte differentiation, inflammation, and cell cycle control. *Immunity* **13**, 199-212.

Shah, K. M. & Young, L. S. (2009). Epstein-Barr virus and carcinogenesis: beyond Burkitt's lymphoma. *Clin Microbiol Infect* **15**, 982-988.

Shannon-Lowe, C., Adland, E., Bell, A. I., Delecluse, H. J., Rickinson, A. B. & Rowe, M. (2009). Features distinguishing Epstein-Barr virus infections of epithelial cells and B cells: viral genome expression, genome maintenance, and genome amplification. *J Virol* **83**, 7749-7760.

Shapiro-Shelef, M. & Calame, K. (2005). Regulation of plasma-cell development. *Nat Rev Immunol* **5**, 230-242.

Shapiro-Shelef, M., Lin, K. I., McHeyzer-Williams, L. J., Liao, J., McHeyzer-Williams, M. G. & Calame, K. (2003). Blimp-1 is required for the formation of immunoglobulin secreting plasma cells and pre-plasma memory B cells. *Immunity* **19**, 607-620.

Shendure, J. & Ji, H. (2008). Next-generation DNA sequencing. *Nat Biotechnol* **26**, 1135-1145.

Shire, K., Ceccarelli, D. F., Avolio-Hunter, T. M. & Frappier, L. (1999). EBP2, a human protein that interacts with sequences of the Epstein-Barr virus nuclear antigen 1 important for plasmid maintenance. *J Virol* **73**, 2587-2595.

Sims, K. S., A., Robertson, E. S., (2010). Regulation of Cellular processes by the Epstein-Barr Virus Nuclear Antigen 3 Family of proteins. In *Epstein-Barr Virus: Latency and Transformation*. Edited by E. S. Robertson: Caister Academic Press.

Sinclair, A. J. (2003). bZIP proteins of human gammaherpesviruses. *J Gen Virol* **84**, 1941-1949.

Sinclair, A. J. & Farrell, P. J. (1992). Epstein-Barr virus transcription factors. *Cell Growth Differ* **3**, 557-563.

Sinclair, A. J., Palmero, I., Peters, G. & Farrell, P. J. (1994). EBNA-2 and EBNA-LP cooperate to cause G0 to G1 transition during immortalization of resting human B lymphocytes by Epstein-Barr virus. *EMBO J* **13**, 3321-3328.

Sjoblom, A., Jansson, A., Yang, W., Lain, S., Nilsson, T. & Rymo, L. (1995). PU box-binding transcription factors and a POU domain protein cooperate in the Epstein-Barr virus (EBV) nuclear antigen 2-induced transactivation of the EBV latent membrane protein 1 promoter. *J Gen Virol* **76 (Pt 11)**, 2679-2692.

Smith, J. M. (1992). Analyzing the mosaic structure of genes. *J Mol Evol* **34**, 126-129.

Soulier, J., Grollet, L., Oksenhendler, E., Cacoub, P., Cazals-Hatem, D., Babinet, P., d'Agay, M. F., Clauvel, J. P., Raphael, M., Degos, L. & et al. (1995). Kaposi's sarcoma-associated herpesvirus-like DNA sequences in multicentric Castleman's disease. *Blood* **86**, 1276-1280.

Souza, T. A., Stollar, B. D., Sullivan, J. L., Luzuriaga, K. & Thorley-Lawson, D. A. (2005). Peripheral B cells latently infected with Epstein-Barr virus display molecular hallmarks of classical antigen-selected memory B cells. *Proc Natl Acad Sci U S A* **102**, 18093-18098.

Speck, S. H. & Strominger, J. L. (1985). Analysis of the transcript encoding the latent Epstein-Barr virus nuclear antigen I: a potentially polycistronic message generated by long-range splicing of several exons. *Proc Natl Acad Sci U S A* **82**, 8305-8309.

Spriggs, M. K., Armitage, R. J., Comeau, M. R., Strockbine, L., Farrah, T., Macduff, B., Ulrich, D., Alderson, M. R., Mullberg, J. & Cohen, J. I. (1996).

The extracellular domain of the Epstein-Barr virus BZLF2 protein binds the HLA-DR beta chain and inhibits antigen presentation. *J Virol* **70**, 5557-5563.

Srivastava, B., Quinn, W. J., 3rd, Hazard, K., Erikson, J. & Allman, D. (2005). Characterization of marginal zone B cell precursors. *J Exp Med* **202**, 1225-1234.

Staskus, K. A., Sun, R., Miller, G., Racz, P., Jaslowski, A., Metroka, C., Brett-Smith, H. & Haase, A. T. (1999). Cellular tropism and viral interleukin-6 expression distinguish human herpesvirus 8 involvement in Kaposi's sarcoma, primary effusion lymphoma, and multicentric Castleman's disease. *J Virol* **73**, 4181-4187.

Stavnezer, J. & Schrader, C. E. (2006). Mismatch repair converts AID-instigated nicks to double-strand breaks for antibody class-switch recombination. *Trends Genet* **22**, 23-28.

Stebbing, J., Bourboulia, D., Johnson, M., Henderson, S., Williams, I., Wilder, N., Tyrer, M., Youle, M., Imami, N., Kobu, T., Kuon, W., Sieper, J., Gotch, F. & Boshoff, C. (2003). Kaposi's sarcoma-associated herpesvirus cytotoxic T lymphocytes recognize and target Darwinian positively selected autologous K1 epitopes. *J Virol* **77**, 4306-4314.

Stein, H. D., G.; Pileri, S.; Said, J.; Mann, R.; Poppema, S.; Jaffe, E.S.; Swerdlow, S.H. (2001). Classical Hodgkin Lymphoma. In *Pathology and Genetics: Tumours of Haematopoietic and Lymphoid Tissues*. Edited by N. L. H. Elaine S. Jaffe, Harald Stein, James W Vardiman: IARC Press, International Agency for REsearch on Cancer (IARC) 69008 Lyon, France.

Storz, P. & Toker, A. (2003). Protein kinase D mediates a stress-induced NF-kappaB activation and survival pathway. *EMBO J* **22**, 109-120.

Strauch, B., Andrews, L. L., Siegel, N. & Miller, G. (1974). Oropharyngeal excretion of Epstein-Barr virus by renal transplant recipients and other patients treated with immunosuppressive drugs. *Lancet* **1**, 234-237.

Sturzl, M., Hohenadl, C., Zietz, C., Castanos-Velez, E., Wunderlich, A., Ascherl, G., Biberfeld, P., Monini, P., Browning, P. J. & Ensoli, B. (1999). Expression of K13/v-FLIP gene of human herpesvirus 8 and apoptosis in Kaposi's sarcoma spindle cells. *J Natl Cancer Inst* **91**, 1725-1733.

Sugden, B. & Warren, N. (1989). A promoter of Epstein-Barr virus that can function during latent infection can be transactivated by EBNA-1, a viral protein required for viral DNA replication during latent infection. *J Virol* **63**, 2644-2649.

Sun, C. C. & Thorley-Lawson, D. A. (2007). Plasma cell-specific transcription factor XBP-1s binds to and transactivates the Epstein-Barr virus BZLF1 promoter. *J Virol* **81**, 13566-13577.

Sun, R., Lin, S. F., Gradoville, L., Yuan, Y., Zhu, F. & Miller, G. (1998). A viral gene that activates lytic cycle expression of Kaposi's sarcoma-associated herpesvirus. *Proc Natl Acad Sci U S A* **95**, 10866-10871.

Sung, N. S., Edwards, R. H., Seillier-Moiseiwitsch, F., Perkins, A. G., Zeng, Y. & Raab-Traub, N. (1998). Epstein-Barr virus strain variation in nasopharyngeal carcinoma from the endemic and non-endemic regions of China. *Int J Cancer* **76**, 207-215.

Sung, N. S., Kenney, S., Gutsch, D. & Pagano, J. S. (1991). EBNA-2 transactivates a lymphoid-specific enhancer in the BamHI C promoter of Epstein-Barr virus. *J Virol* **65**, 2164-2169.

Szakonyi, G., Guthridge, J. M., Li, D., Young, K., Holers, V. M. & Chen, X. S. (2001). Structure of complement receptor 2 in complex with its C3d ligand. *Science* **292**, 1725-1728.

Ta, V. T., Nagaoka, H., Catalan, N., Durandy, A., Fischer, A., Imai, K., Nonoyama, S., Tashiro, J., Ikegawa, M., Ito, S., Kinoshita, K., Muramatsu, M. & Honjo, T. (2003). AID mutant analyses indicate requirement for class-switch-specific cofactors. *Nat Immunol* **4**, 843-848.

Takada, K. (1984). Cross-linking of cell surface immunoglobulins induces Epstein-Barr virus in Burkitt lymphoma lines. *Int J Cancer* **33**, 27-32.

Takada, K., Horinouchi, K., Ono, Y., Aya, T., Osato, T., Takahashi, M. & Hayasaka, S. (1991). An Epstein-Barr virus-producer line Akata: establishment of the cell line and analysis of viral DNA. *Virus Genes* **5**, 147-156.

Takada, K. & Ono, Y. (1989). Synchronous and sequential activation of latently infected Epstein-Barr virus genomes. *J Virol* **63**, 445-449.

Takada, K., Shimizu, N., Sakuma, S. & Ono, Y. (1986). trans activation of the latent Epstein-Barr virus (EBV) genome after transfection of the EBV DNA fragment. *J Virol* **57**, 1016-1022.

Takahashi, Y., Cerasoli, D. M., Dal Porto, J. M., Shimoda, M., Freund, R., Fang, W., Telander, D. G., Malvey, E. N., Mueller, D. L., Behrens, T. W. & Kelsoe, G. (1999). Relaxed negative selection in germinal centers and impaired affinity maturation in bcl-xL transgenic mice. *J Exp Med* **190**, 399-410.

Tanner, J., Weis, J., Fearon, D., Whang, Y. & Kieff, E. (1987). Epstein-Barr virus gp350/220 binding to the B lymphocyte C3d receptor mediates adsorption, capping, and endocytosis. *Cell* **50**, 203-213.

Tao, Q. & Robertson, K. D. (2003). Stealth technology: how Epstein-Barr virus utilizes DNA methylation to cloak itself from immune detection. *Clin Immunol* **109**, 53-63.

Tao, Q., Young, L. S., Woodman, C. B. & Murray, P. G. (2006). Epstein-Barr virus (EBV) and its associated human cancers--genetics, epigenetics, pathobiology and novel therapeutics. *Front Biosci* **11**, 2672-2713.

Taylor, G. M., Raghuwanshi, S. K., Rowe, D. T., Wadowsky, R. M. & Rosendorff, A. (2011). Endoplasmic reticulum stress causes EBV lytic replication. *Blood* **118**, 5528-5539.

Thorley-Lawson, D. A. (2001). Epstein-Barr virus: exploiting the immune system. *Nat Rev Immunol* **1**, 75-82.

Thorley-Lawson, D. A. (2005). EBV the prototypical human tumor virus--just how bad is it? *J Allergy Clin Immunol* **116**, 251-261; quiz 262.

Thorley-Lawson, D. A. & Babcock, G. J. (1999). A model for persistent infection with Epstein-Barr virus: the stealth virus of human B cells. *Life Sci* **65**, 1433-1453.

Thorley-Lawson, D. A. & Gross, A. (2004). Persistence of the Epstein-Barr virus and the origins of associated lymphomas. *N Engl J Med* **350**, 1328-1337.

Thuerauf, D. J., Morrison, L. & Glembotski, C. C. (2004). Opposing roles for ATF6alpha and ATF6beta in endoplasmic reticulum stress response gene induction. *J Biol Chem* **279**, 21078-21084.

Tiegs, S. L., Russell, D. M. & Nemazee, D. (2011). Receptor editing in self-reactive bone marrow B cells. The Journal of Experimental Medicine. 1993. 177: 1009-1020. *J Immunol* **186**, 1313-1324.

Tierney, R., Kirby, H., Nagra, J., Rickinson, A. & Bell, A. (2000a). The Epstein-Barr virus promoter initiating B-cell transformation is activated by RFX

proteins and the B-cell-specific activator protein BSAP/Pax5. *J Virol* **74**, 10458-10467.

Tierney, R., Nagra, J., Hutchings, I., Shannon-Lowe, C., Altmann, M., Hammerschmidt, W., Rickinson, A. & Bell, A. (2007). Epstein-Barr virus exploits BSAP/Pax5 to achieve the B-cell specificity of its growth-transforming program. *J Virol* **81**, 10092-10100.

Tierney, R. J., Kao, K. Y., Nagra, J. K. & Rickinson, A. B. (2011). Epstein-Barr virus BamHI W repeat number limits EBNA2/EBNA-LP coexpression in newly infected B cells and the efficiency of B-cell transformation: a rationale for the multiple W repeats in wild-type virus strains. *J Virol* **85**, 12362-12375.

Tierney, R. J., Kirby, H. E., Nagra, J. K., Desmond, J., Bell, A. I. & Rickinson, A. B. (2000b). Methylation of transcription factor binding sites in the Epstein-Barr virus latent cycle promoter Wp coincides with promoter down-regulation during virus-induced B-cell transformation. *J Virol* **74**, 10468-10479.

Todd, D. J., McHeyzer-Williams, L. J., Kowal, C., Lee, A. H., Volpe, B. T., Diamond, B., McHeyzer-Williams, M. G. & Glimcher, L. H. (2009). XBP1 governs late events in plasma cell differentiation and is not required for antigen-specific memory B cell development. *J Exp Med* **206**, 2151-2159.

Toellner, K. M., Gulbranson-Judge, A., Taylor, D. R., Sze, D. M. & MacLennan, I. C. (1996). Immunoglobulin switch transcript production in vivo related to the site and time of antigen-specific B cell activation. *J Exp Med* **183**, 2303-2312.

Tomkinson, B. & Kieff, E. (1992). Use of second-site homologous recombination to demonstrate that Epstein-Barr virus nuclear protein 3B is not important for lymphocyte infection or growth transformation in vitro. *J Virol* **66**, 2893-2903.

Tomkinson, B., Robertson, E. & Kieff, E. (1993). Epstein-Barr virus nuclear proteins EBNA-3A and EBNA-3C are essential for B-lymphocyte growth transformation. *J Virol* **67**, 2014-2025.

Toussirot, E. & Roudier, J. (2008). Epstein-Barr virus in autoimmune diseases. *Best Pract Res Clin Rheumatol* **22**, 883-896.

Toyama, H., Okada, S., Hatano, M., Takahashi, Y., Takeda, N., Ichii, H., Takemori, T., Kuroda, Y. & Tokuhisa, T. (2002). Memory B cells without somatic hypermutation are generated from Bcl6-deficient B cells. *Immunity* **17**, 329-339.

Triantos, D., Leao, J. C., Porter, S. R., Scully, C. M. & Teo, C. G. (1998). Tissue distribution of Epstein-Barr virus genotypes in hosts coinfected by HIV. *AIDS* **12**, 2141-2146.

Trivedi, P., Takazawa, K., Zompetta, C., Cuomo, L., Anastasiadou, E., Carbone, A., Uccini, S., Belardelli, F., Takada, K., Frati, L. & Faggioni, A. (2004). Infection of HHV-8+ primary effusion lymphoma cells with a recombinant Epstein-Barr virus leads to restricted EBV latency, altered phenotype, and increased tumorigenicity without affecting TCL1 expression. *Blood* **103**, 313-316.

Tsai, K., Thikmyanova, N., Wojcechowskyj, J. A., Delecluse, H. J. & Lieberman, P. M. (2011). EBV tegument protein BNRF1 disrupts DAXX-ATRX to activate viral early gene transcription. *PLoS Pathog* **7**, e1002376.

Turcatti, G., Romieu, A., Fedurco, M. & Tairi, A. P. (2008). A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic Acids Res* **36**, e25.

Turner, C. A., Jr., Mack, D. H. & Davis, M. M. (1994). Blimp-1, a novel zinc finger-containing protein that can drive the maturation of B lymphocytes into immunoglobulin-secreting cells. *Cell* **77**, 297-306.

Tyler, S. D., Peters, G. A., Grose, C., Severini, A., Gray, M. J., Upton, C. & Tipples, G. A. (2007). Genomic cartography of varicella-zoster virus: a complete genome-based analysis of strain variability with implications for attenuation and phenotypic differences. *Virology* **359**, 447-458.

Urano, F., Wang, X., Bertolotti, A., Zhang, Y., Chung, P., Harding, H. P. & Ron, D. (2000). Coupling of stress in the ER to activation of JNK protein kinases by transmembrane protein kinase IRE1. *Science* **287**, 664-666.

Valencia, S. M. & Hutt-Fletcher, L. M. (2012). Important but differential roles for actin in trafficking of Epstein-Barr virus in B cells and epithelial cells. *J Virol* **86**, 2-10.

Vereide, D. T. & Sugden, B. (2011). Lymphomas differ in their dependence on Epstein-Barr virus. *Blood* **117**, 1977-1985.

Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E. & Andino, R. (2006). Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* **439**, 344-348.

Walling, D. M., Brown, A. L., Etienne, W., Keitel, W. A. & Ling, P. D. (2003). Multiple Epstein-Barr virus infections in healthy individuals. *J Virol* **77**, 6546-6550.

Waltzer, L., Perricaudet, M., Sergeant, A. & Manet, E. (1996). Epstein-Barr virus EBNA3A and EBNA3C proteins both repress RBP-J kappa-EBNA2-activated transcription by inhibiting the binding of RBP-J kappa to DNA. *J Virol* **70**, 5909-5915.

Wang, D., Liebowitz, D. & Kieff, E. (1985). An EBV membrane protein expressed in immortalized lymphocytes transforms established rodent cells. *Cell* **43**, 831-840.

Wang, F., Gregory, C., Sample, C., Rowe, M., Liebowitz, D., Murray, R., Rickinson, A. & Kieff, E. (1990a). Epstein-Barr virus latent membrane protein (LMP1) and nuclear proteins 2 and 3C are effectors of phenotypic changes in B lymphocytes: EBNA-2 and LMP1 cooperatively induce CD23. *J Virol* **64**, 2309-2318.

Wang, F., Gregory, C. D., Rowe, M., Rickinson, A. B., Wang, D., Birkenbach, M., Kikutani, H., Kishimoto, T. & Kieff, E. (1987). Epstein-Barr virus nuclear antigen 2 specifically induces expression of the B-cell activation antigen CD23. *Proc Natl Acad Sci U S A* **84**, 3452-3456.

Wang, F., Tsang, S. F., Kurilla, M. G., Cohen, J. I. & Kieff, E. (1990b). Epstein-Barr virus nuclear antigen 2 transactivates latent membrane protein LMP1. *J Virol* **64**, 3407-3416.

Wang, J. T., Sheeng, T. S., Su, I. J., Chen, J. Y. & Chen, M. R. (2003). EBNA-1 sequence variations reflect active EBV replication and disease status or quiescent latency in lymphocytes. *J Med Virol* **69**, 417-425.

Wang, Q. J. (2006). PKD at the crossroads of DAG and PKC signaling. *Trends Pharmacol Sci* **27**, 317-323.

Wang, X. & Hutt-Fletcher, L. M. (1998). Epstein-Barr virus lacking glycoprotein gp42 can bind to B cells but is not able to infect. *J Virol* **72**, 158-163.

Wang, Y. C., Huang, J. M. & Montalvo, E. A. (1997). Characterization of proteins binding to the ZII element in the Epstein-Barr virus BZLF1 promoter: transactivation by ATF1. *Virology* **227**, 323-330.

Wasserman, R., Li, Y. S., Shinton, S. A., Carmack, C. E., Manser, T., Wiest, D. L., Hayakawa, K. & Hardy, R. R. (1998). A novel mechanism for B cell repertoire maturation based on response by B cell precursors to pre-B receptor assembly. *J Exp Med* **187**, 259-264.

Weis, J. J., Tedder, T. F. & Fearon, D. T. (1984). Identification of a 145,000 Mr membrane protein as the C3d receptor (CR2) of human B lymphocytes. *Proc Natl Acad Sci U S A* **81**, 881-885.

Weiss, L. M. (2000). Epstein-Barr virus and Hodgkin's disease. *Curr Oncol Rep* **2**, 199-204.

Weller, S., Braun, M. C., Tan, B. K., Rosenwald, A., Cordier, C., Conley, M. E., Plebani, A., Kumararatne, D. S., Bonnet, D., Tournilhac, O., Tchernia, G., Steiniger, B., Staudt, L. M., Casanova, J. L., Reynaud, C. A. & Weill, J. C. (2004). Human blood IgM "memory" B cells are circulating splenic marginal zone B cells harboring a prediversified immunoglobulin repertoire. *Blood* **104**, 3647-3654.

Werner, M., Ernberg, I., Zou, J., Almqvist, J. & Aurell, E. (2007). Epstein-Barr virus latency switch in human B-cells: a physico-chemical model. *BMC Syst Biol* **1**, 40.

West, M. J., Webb, H. M., Sinclair, A. J. & Woolfson, D. N. (2004). Biophysical and mutational analysis of the putative bZIP domain of Epstein-Barr virus EBNA 3C. *J Virol* **78**, 9431-9445.

White, R. E., Ramer, P. C., Naresh, K. N., Meixlsperger, S., Pinaud, L., Rooney, C., Savoldo, B., Coutinho, R., Bodor, C., Gribben, J., Ibrahim, H. A., Bower, M., Nourse, J. P., Gandhi, M. K., Middeldorp, J., Cader, F. Z., Murray, P., Munz, C. & Allday, M. J. (2012). EBNA3B-deficient EBV promotes B cell lymphomagenesis in humanized mice and is found in human tumors. *J Clin Invest* **122**, 1487-1502.

Wilson, S. J., Tsao, E. H., Webb, B. L., Ye, H., Dalton-Griffin, L., Tsantoulas, C., Gale, C. V., Du, M. Q., Whitehouse, A. & Kellam, P. (2007). X box binding protein XBP-1s transactivates the Kaposi's sarcoma-associated herpesvirus (KSHV) ORF50 promoter, linking plasma cell differentiation to KSHV reactivation from latency. *J Virol* **81**, 13578-13586.

Woisetschlaeger, M., Jin, X. W., Yandava, C. N., Furmanski, L. A., Strominger, J. L. & Speck, S. H. (1991). Role for the Epstein-Barr virus nuclear antigen 2 in viral promoter switching during initial stages of infection. *Proc Natl Acad Sci U S A* **88**, 3942-3946.

Woisetschlaeger, M., Yandava, C. N., Furmanski, L. A., Strominger, J. L. & Speck, S. H. (1990). Promoter switching in Epstein-Barr virus during the initial stages of infection of B lymphocytes. *Proc Natl Acad Sci U S A* **87**, 1725-1729.

Wu, H., Kapoor, P. & Frappier, L. (2002). Separation of the DNA replication, segregation, and transcriptional activation functions of Epstein-Barr nuclear antigen 1. *J Virol* **76**, 2480-2490.

Wysocka, J. & Herr, W. (2003). The herpes simplex virus VP16-induced complex: the makings of a regulatory switch. *Trends Biochem Sci* **28**, 294-304.

Xie, J., Ajibade, A. O., Ye, F., Kuhne, K. & Gao, S. J. (2008). Reactivation of Kaposi's sarcoma-associated herpesvirus from latency requires MEK/ERK,

JNK and p38 multiple mitogen-activated protein kinase pathways. *Virology* **371**, 139-154.

Xu, D., Coleman, T., Zhang, J., Fagot, A., Kotalik, C., Zhao, L., Trivedi, P., Jones, C. & Zhang, L. (2007). Epstein-Barr virus inhibits Kaposi's sarcoma-associated herpesvirus lytic replication in primary effusion lymphomas. *J Virol* **81**, 6068-6078.

Yalamanchili, R., Tong, X., Grossman, S., Johannsen, E., Mosialos, G. & Kieff, E. (1994). Genetic and biochemical evidence that EBNA 2 interaction with a 63-kDa cellular GTG-binding protein is essential for B lymphocyte growth transformation by EBV. *Virology* **204**, 634-641.

Yamamoto, K., Sato, T., Matsui, T., Sato, M., Okada, T., Yoshida, H., Harada, A. & Mori, K. (2007). Transcriptional induction of mammalian ER quality control proteins is mediated by single or combined action of ATF6alpha and XBP1. *Dev Cell* **13**, 365-376.

Yan, W., Frank, C. L., Korth, M. J., Sopher, B. L., Novoa, I., Ron, D. & Katze, M. G. (2002). Control of PERK eIF2alpha kinase activity by the endoplasmic reticulum stress-induced molecular chaperone P58IPK. *Proc Natl Acad Sci U S A* **99**, 15920-15925.

Yao, Q. Y., Ogan, P., Rowe, M., Wood, M. & Rickinson, A. B. (1989). Epstein-Barr virus-infected B cells persist in the circulation of acyclovir-treated virus carriers. *Int J Cancer* **43**, 67-71.

Yao, Q. Y., Rickinson, A. B. & Epstein, M. A. (1985). Oropharyngeal shedding of infectious Epstein-Barr virus in healthy virus-immune donors. A prospective study. *Chin Med J (Engl)* **98**, 191-196.

Yao, Q. Y., Tierney, R. J., Croom-Carter, D., Dukers, D., Cooper, G. M., Ellis, C. J., Rowe, M. & Rickinson, A. B. (1996). Frequency of multiple Epstein-Barr virus infections in T-cell-immunocompromised individuals. *J Virol* **70**, 4884-4894.

Yaswen, L. R., Stephens, E. B., Davenport, L. C. & Hutt-Fletcher, L. M. (1993). Epstein-Barr virus glycoprotein gp85 associates with the BKRF2 gene product and is incompletely processed as a recombinant protein. *Virology* **195**, 387-396.

Yates, J., Warren, N., Reisman, D. & Sugden, B. (1984). A cis-acting element from the Epstein-Barr viral genome that permits stable replication of recombinant plasmids in latently infected cells. *Proc Natl Acad Sci U S A* **81**, 3806-3810.

Yates, J. L., Camiolo, S. M. & Bashaw, J. M. (2000). The minimal replicator of Epstein-Barr virus oriP. *J Virol* **74**, 4512-4522.

Yates, J. L., Warren, N. & Sugden, B. (1985). Stable replication of plasmids derived from Epstein-Barr virus in various mammalian cells. *Nature* **313**, 812-815.

Ye, J., Rawson, R. B., Komuro, R., Chen, X., Dave, U. P., Prywes, R., Brown, M. S. & Goldstein, J. L. (2000). ER stress induces cleavage of membrane-bound ATF6 by the same proteases that process SREBPs. *Mol Cell* **6**, 1355-1364.

Yin, Y., Manoury, B. & Fahraeus, R. (2003). Self-inhibition of synthesis and antigen presentation by Epstein-Barr virus-encoded EBNA1. *Science* **301**, 1371-1374.

Yoneda, T., Imaizumi, K., Oono, K., Yui, D., Gomi, F., Katayama, T. & Tohyama, M. (2001). Activation of caspase-12, an endoplastic reticulum (ER) resident caspase, through tumor necrosis factor receptor-associated factor 2-

dependent mechanism in response to the ER stress. *J Biol Chem* **276**, 13935-13940.

Yoo, L. & Speck, S. H. (2000). Determining the role of the Epstein-Barr virus Cp EBNA2-dependent enhancer during the establishment of latency by using mutant and wild-type viruses recovered from cottontop marmoset lymphoblastoid cell lines. *J Virol* **74**, 11115-11120.

Yoshida, H., Okada, T., Haze, K., Yanagi, H., Yura, T., Negishi, M. & Mori, K. (2000). ATF6 activated by proteolysis binds in the presence of NF-Y (CBF) directly to the cis-acting element responsible for the mammalian unfolded protein response. *Mol Cell Biol* **20**, 6755-6767.

Yoshida, H., Oku, M., Suzuki, M. & Mori, K. (2006). pXBP1(U) encoded in XBP1 pre-mRNA negatively regulates unfolded protein response activator pXBP1(S) in mammalian ER stress response. *J Cell Biol* **172**, 565-575.

Younes, B. S., Ament, M. E., McDiarmid, S. V., Martin, M. G. & Vargas, J. H. (1999). The involvement of the gastrointestinal tract in posttransplant lymphoproliferative disease in pediatric liver transplantation. *J Pediatr Gastroenterol Nutr* **28**, 380-385.

Young, L. S. & Murray, P. G. (2003). Epstein-Barr virus and oncogenesis: from latent genes to tumours. *Oncogene* **22**, 5108-5121.

Young, L. S. & Rickinson, A. B. (2004). Epstein-Barr virus: 40 years on. *Nat Rev Cancer* **4**, 757-768.

Yu, F., Feng, J., Harada, J. N., Chanda, S. K., Kenney, S. C. & Sun, R. (2007a). B cell terminal differentiation factor XBP-1 induces reactivation of Kaposi's sarcoma-associated herpesvirus. *FEBS Lett* **581**, 3485-3488.

Yu, X., Wang, Z. & Mertz, J. E. (2007b). ZEB1 regulates the latent-lytic switch in infection by Epstein-Barr virus. *PLoS Pathog* **3**, e194.

Yuan, J., Cahir-McFarland, E., Zhao, B. & Kieff, E. (2006). Virus and cell RNAs expressed during Epstein-Barr virus replication. *J Virol* **80**, 2548-2565.

Yue, W., Davenport, M. G., Shackelford, J. & Pagano, J. S. (2004). Mitosis-specific hyperphosphorylation of Epstein-Barr virus nuclear antigen 2 suppresses its function. *J Virol* **78**, 3542-3552.

Zagordi, O., Klein, R., Daumer, M. & Beerenwinkel, N. (2010). Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies. *Nucleic Acids Res* **38**, 7400-7409.

Zalani, S., Holley-Guthrie, E. & Kenney, S. (1996). Epstein-Barr viral latency is disrupted by the immediate-early BRLF1 protein through a cell-specific mechanism. *Proc Natl Acad Sci U S A* **93**, 9194-9199.

Zandvoort, A. & Timens, W. (2002). The dual function of the splenic marginal zone: essential for initiation of anti-TI-2 responses but also vital in the general first-line defense against blood-borne antigens. *Clin Exp Immunol* **130**, 4-11.

Zhong, W., Wang, H., Herndier, B. & Ganem, D. (1996). Restricted expression of Kaposi sarcoma-associated herpesvirus (human herpesvirus 8) genes in Kaposi sarcoma. *Proc Natl Acad Sci U S A* **93**, 6641-6646.

Zimber, U., Adldinger, H. K., Lenoir, G. M., Vuillaume, M., Knebel-Doeberitz, M. V., Laux, G., Desgranges, C., Wittmann, P., Freese, U. K., Schneider, U. & et al. (1986). Geographical prevalence of two types of Epstein-Barr virus. *Virology* **154**, 56-66.

Zimber-Strobl, U., Kempkes, B., Marschall, G., Zeidler, R., Van Kooten, C., Banchereau, J., Bornkamm, G. W. & Hammerschmidt, W. (1996). Epstein-Barr virus latent membrane protein (LMP1) is not sufficient to

maintain proliferation of B cells but both it and activated CD40 can prolong their survival. *EMBO J* **15**, 7070-7078.

Zimber-Strobl, U., Suentzenich, K. O., Laux, G., Eick, D., Cordier, M., Calender, A., Billaud, M., Lenoir, G. M. & Bornkamm, G. W. (1991). Epstein-Barr virus nuclear antigen 2 activates transcription of the terminal protein gene. *J Virol* **65**, 415-423.

Zinszner, H., Kuroda, M., Wang, X., Batchvarova, N., Lightfoot, R. T., Remotti, H., Stevens, J. L. & Ron, D. (1998). CHOP is implicated in programmed cell death in response to impaired function of the endoplasmic reticulum. *Genes Dev* **12**, 982-995.

Zong, J., Ciufo, D. M., Viscidi, R., Alagiozoglou, L., Tyring, S., Rady, P., Orenstein, J., Boto, W., Kalumbuja, H., Romano, N., Melbye, M., Kang, G. H., Boshoff, C. & Hayward, G. S. (2002). Genotypic analysis at multiple loci across Kaposi's sarcoma herpesvirus (KSHV) DNA molecules: clustering patterns, novel variants and chimerism. *J Clin Virol* **23**, 119-148.

Zong, J. C., Ciufo, D. M., Alcendor, D. J., Wan, X., Nicholas, J., Browning, P. J., Rady, P. L., Tyring, S. K., Orenstein, J. M., Rabkin, C. S., Su, I. J., Powell, K. F., Croxson, M., Foreman, K. E., Nickoloff, B. J., Alkan, S. & Hayward, G. S. (1999). High-level variability in the ORF-K1 membrane protein gene at the left end of the Kaposi's sarcoma-associated herpesvirus genome defines four major virus subtypes and multiple variants or clades in different human populations. *J Virol* **73**, 4156-4170.

Zugaza, J. L., Sinnett-Smith, J., Van Lint, J. & Rozengurt, E. (1996). Protein kinase D (PKD) activation in intact cells through a protein kinase C-dependent signal transduction pathway. *EMBO J* **15**, 6220-6230.

zur Hausen, H., Bornkamm, G. W., Schmidt, R. & Hecker, E. (1979). Tumor initiators and promoters in the induction of Epstein-Barr virus. *Proc Natl Acad Sci U S A* **76**, 782-785.

Journal of General Virology (2011), **92**, 421–431 DOI 10.1099/vir.0.025494-0

X-box binding protein 1 induces the expression of the lytic cycle transactivator of Kaposi's sarcoma-associated herpesvirus but not Epstein–Barr virus in co-infected primary effusion lymphoma

Imogen Yi-Chun Lai,[1] Paul J. Farrell[2] and Paul Kellam[1,3]

Correspondence
Paul Kellam
pk5@sanger.ac.uk

[1]University College London, MRC Centre for Molecular Virology, Department of Infection, Division of Infection and Immunity, Windeyer Institute of Medical Science, 46 Cleveland Street, London W1T 4JF, UK

[2]Section of Virology, Imperial College Faculty of Medicine, St Mary's Campus, Norfolk Place, London W2 1PG, UK

[3]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Received 19 July 2010
Accepted 22 October 2010

Cells of primary effusion lymphoma (PEL), a B-cell non-Hodgkin's lymphoma, are latently infected by Kaposi's sarcoma-associated herpesvirus (KSHV), with about 80 % of PEL also co-infected with Epstein–Barr virus (EBV). Both viruses can be reactivated into their lytic replication cycle in PEL by chemical inducers. However, simultaneous activation of both lytic cascades leads to mutual lytic cycle co-repression. The plasma cell-differentiation factor X-box binding protein 1 (XBP-1) transactivates the KSHV immediate–early promoter leading to the production of the replication and transcription activator protein (RTA), and reactivation of KSHV from latency. XBP-1 has been reported to act similarly on the EBV immediate–early promoter Zp, leading to the production of the lytic-cycle transactivator protein BZLF1. Here we show that activated B-cell terminal-differentiation transcription factor X-box binding protein 1 (XBP-1s) does not induce EBV BZLF1 and BRLF1 expression in PEL and BL cell lines, despite inducing lytic reactivation of KSHV in PEL. We show that XBP-1s transactivates the KSHV RTA promoter but does not transactivate the EBV BZLF1 promoter in non-B-cells by using a luciferase assay. Co-expression of activated protein kinase D, which can phosphorylate and inactivate class II histone deacetylases (HDACs), does not rescue XBP-1 activity on Zp nor does it induce BZLF1 and BRLF1 expression in PEL. Finally, chemical inducers of KSHV and EBV lytic replication in PEL, including HDAC inhibitors, do not lead to XBP-1 activation. We conclude that XBP-1 specifically reactivates the KSHV lytic cycle in dually infected PELs.

INTRODUCTION

The human gammaherpesviruses Epstein–Barr virus (EBV) and Kaposi's sarcoma-associated herpesvirus (KSHV) are associated with B-cell lymphomas and tumours of epithelial and endothelial origin, respectively. EBV and KSHV co-infection occurs in approximately 80 % of the non-Hodgkin's B-cell lymphoma, primary effusion lymphoma (PEL); the remainder being infected by KSHV alone. KSHV and EBV dually infected PEL have a subtly different pattern of B-cell gene expression compared with KSHV singly infected PEL (Fan *et al.*, 2005). In addition, the presence of EBV appears to potentiate the tumourigenicity of dually infected PELs in SCID mice (Trivedi *et al.*, 2004). How dually infected PEL arise is not understood, however, both KSHV and EBV latently infect PEL and normal B-cells.

EBV establishes a latent infection in memory B-cells *in vivo* (Babcock *et al.*, 1998, 1999; Souza *et al.*, 2005; Thorley-Lawson, 2001). A model for EBV infection suggests that, following saliva transmission, EBV infects IgD-positive antigen-naïve B-cells in the nasopharyngeal lymphoid tissue (Joseph *et al.*, 2000) and induces B-cell proliferation through the expression of virus latent growth programme genes (Thorley-Lawson & Gross, 2004). The EBV-positive lymphoblasts subsequently express the default/latency II gene-expression programme (Thorley-Lawson, 2001; Thorley-Lawson & Gross, 2004), finally entering the memory B-cell repertoire, where the virus silences its gene expression allowing lifelong persistence (Gires *et al.*, 1997; Panagopoulos *et al.*, 2004). The type of B-cell and the mechanism of latent colonization by KSHV are not known.

Horizontal virus transmission requires either EBV and/or KSHV to reactivate from latency with infectious EBV and KSHV being transmitted by saliva. The molecular events that lead to virus reactivation have been studied extensively.

For EBV the reactivation of the lytic cycle is mediated by two viral proteins, BZLF1 (ZTA) and BRLF1 (RTA) (Amon *et al.*, 2004; Chang & Liu, 2000; Yuan *et al.*, 2006). BZLF1 is the key immediate–early protein in EBV (Bryant & Farrell, 2002; Countryman & Miller, 1985), with B-cell receptor (BCR) cross-linking able to activate the BZLF1 promoter (Zp) but not the BRLF1 promoter (Rp). Induction of Zp is therefore the first event of EBV lytic reactivation (Amon *et al.*, 2004), and BZLF1 expression alone is sufficient to initiate the entire EBV lytic cycle (Countryman & Miller, 1985; Rooney *et al.*, 1989). For KSHV, the virus immediate–early replication and transcription activator protein (RTA) is both necessary and sufficient to activate the virus lytic cycle. RTA is able to transactivate its own promoter and those of many viral genes, leading to the KSHV lytic gene expression cascade (Lukac *et al.*, 1998).

The authentic cellular mediators that lead to the induction of EBV BZLF1 (ZTA) and KSHV RTA (K-RTA) are less clear. Previously, we and others showed that the activated B-cell terminal-differentiation transcription factor X-box binding protein 1 (XBP-1s) is sufficient to initiate the lytic cycle of KSHV in PEL cell lines (Dalton-Griffin *et al.*, 2009; Sun & Thorley-Lawson, 2007; Wilson *et al.*, 2007). XBP-1s binds to and transactivates the promoter of KSHV RTA. XBP-1s expression is essential for terminal differentiation of plasma cells (Reimold *et al.*, 2001). XBP-1 retains a 26 nt intron when expressed (XBP-1u, unspliced) that results in a frame shift preventing the translation of an active transcription factor. Under endoplasmic reticulum (ER) stress or B-cell terminal differentiation, the 26 nt intron of XBP-1u mRNA is removed to allow the translation of the active form, XBP-1s (XBP-1 spliced) (Calfon *et al.*, 2002). EBV lytic replication is induced in plasma cells (Laichalk & Thorley-Lawson, 2005); however, EBV infection of multiple myeloma cell lines does not result in lytic EBV replication even though XBP-1s is abundantly expressed in these cells (Anastasiadou *et al.*, 2009). Interestingly, other studies have also suggested that XBP-1s is able to bind to and transactivate the EBV BZLF1 promoter (Bhende *et al.*, 2007; Sun & Thorley-Lawson, 2007). Although XBP-1 can be transiently activated following BCR cross-linking (McDonald *et al.*, 2010), XBP-1s alone may not be sufficient to induce expression of EBV ZTA as protein kinase D (PKD) may also be required (Bhende *et al.*, 2007).

In dually infected PEL a selective lytic switch has been proposed that leads to the induction of one virus over the other (Miller *et al.*, 1997). Indeed K-RTA can only induce the KSHV lytic cycle and BZLF1 can only induce the EBV lytic cycle in PEL. The selective switch can be maintained by the induced virus blocking the lytic replication of the other latent virus (Xu *et al.*, 2007). If both K-RTA and BZLF1 are expressed at the same time, physical interactions promote mutual inhibition of the two transcription factors (Jiang *et al.*, 2008). Here we investigate the mechanism that underlies the initial induction of K-RTA and BZLF1 in dually infected PEL (Miller *et al.*, 1997). We show that XBP-1s alone induces the KSHV lytic cycle but does not induce either EBV BZLF1 or BRLF1 in PEL cell lines; it also does not induce either of the EBV immediately–early proteins in EBV-positive Burkitt's lymphoma cell lines (BLs). Additionally, we show that XBP-1s does not activate the BZLF1 promoter of EBV in HEK 293T and HeLa cells. Overexpressing XBP-1s in both PEL and BLs does not increase the level of BZLF1 and BRLF1 mRNA. These results suggest that XBP-1s activates the KSHV lytic cycle alone in EBV co-infected PEL cells.

RESULTS

XBP-1s overexpression in PEL cell lines results in KSHV RTA expression but not EBV BZLF1 expression

In order to examine the ability of XBP-1s to induce the EBV lytic cycle, we used lentiviral vectors for the transient delivery of XBP-1s into different PEL cell lines. The effect of XBP-1s overexpression on both KSHV and EBV were evaluated by Western blotting against the KSHV immediate–early gene K-RTA and the EBV immediate–early gene BZLF1 (ZTA), respectively. Overexpressing XBP-1s in JSC-1 cells, a KSHV and EBV co-infected PEL, induced K-RTA expression. However, no EBV BZLF1 protein expression was detected, despite the ability of *O*-tetradecanoyl-phorbol 13-acetate (TPA) to induce the expression of K-RTA and BZLF1 in JSC-1 simultaneously (Fig. 1a).

It is possible that EBV in PEL cell lines is in some way defective or non-responsive to XBP-1s. Therefore, we repeated the experiment with BC3 clone 6 (BC3 cl6) and CRO6 clone 2 (CRO6 cl2); two cell lines that were produced by superinfecting KSHV single-positive PEL cell lines with replication-competent EBV–GFP virus (Xu *et al.*, 2007). Western blot analysis showed identical results with the induction of K-RTA but no expression of BZLF1 (Fig. 1b, c) after overexpressing XBP-1s in these cells. TPA treatment again resulted in induction of K-RTA and BZLF1. In all cases, 60–80 % of PEL cells were transduced by the XBP-1s-expressing lentivirus (data not shown). Expression of XBP-1s was confirmed by using Western blot analysis (Fig. 1a, b). Therefore, XBP-1s induces K-RTA expression in PELs, but does not induce BZLF1 expression.

Overexpression of XBP-1s in BL cells does not induce EBV BZLF1 expression

It is possible that the presence of KSHV in PEL affects the ability of XBP-1s to induce BZLF1 expression. We therefore expressed XBP-1s in BLs, which are only EBV positive. Using BCR cross-linking we could induce BZLF1 expression (Fig. 2). However, overexpressing XBP-1s in two different BL lines, Mutu and Akata, did not induce BZLF1 expression despite good transduction efficiencies of 10–25 %. This suggests that XBP-1s does not transactivate the BZLF1 promoter in B-cell lymphomas regardless of the presence of KSHV.

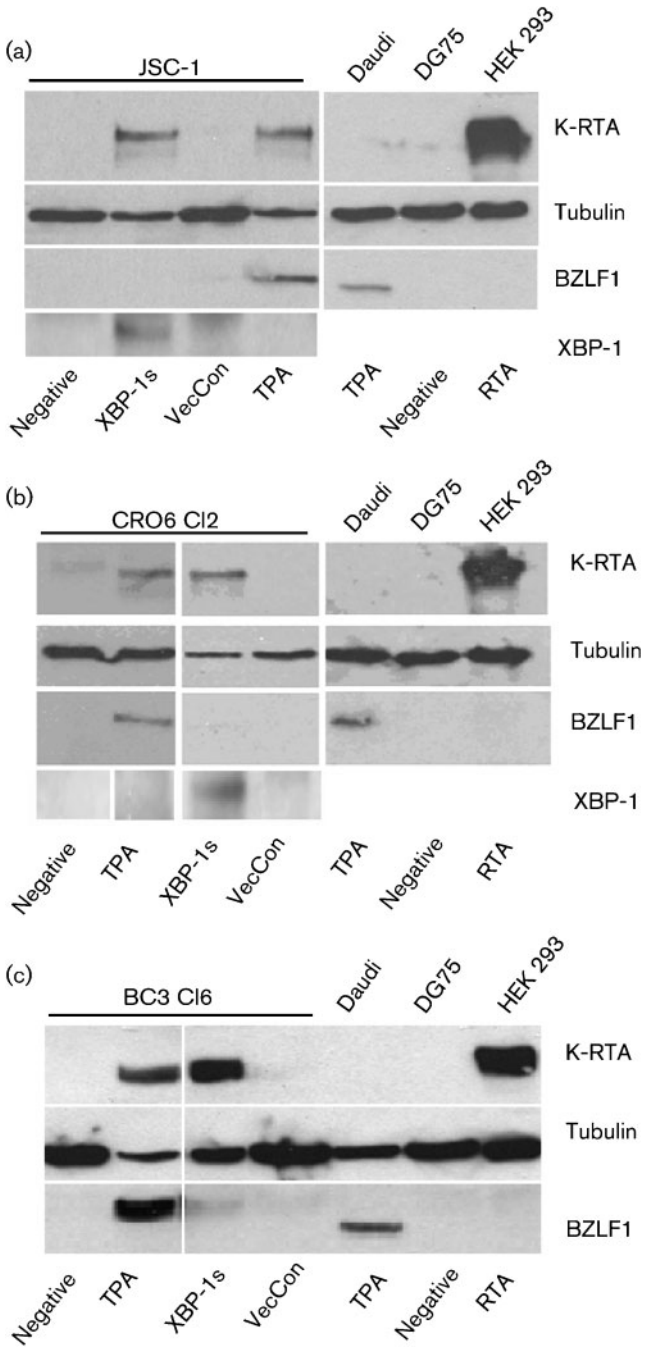


Fig. 1. XBP-1s does not activate EBV lytic gene expression in PEL cell lines. (a) JSC-1, a KSHV and EBV double-positive PEL cell line, expressed K-RTA and BZLF1 after treatment with TPA, but only K-RTA was expressed following XBP-1s transduction. (b) BC3 cl6, and (c) CRO6 cl2 are EBV-superinfected KSHV-positive PEL cell lines and showed a similar pattern with XBP-1s transduction only resulting in K-RTA expression. TPA was able to induce both K-RTA and BZLF1. The Daudi cell line treated with TPA and K-RTA-transfected HEK 293T cells acted as positive controls for the BZLF1 and K-RTA antibody staining, respectively. XBP-1s expression was detected using an XBP-1s specific antibody. VecCon, vector control.

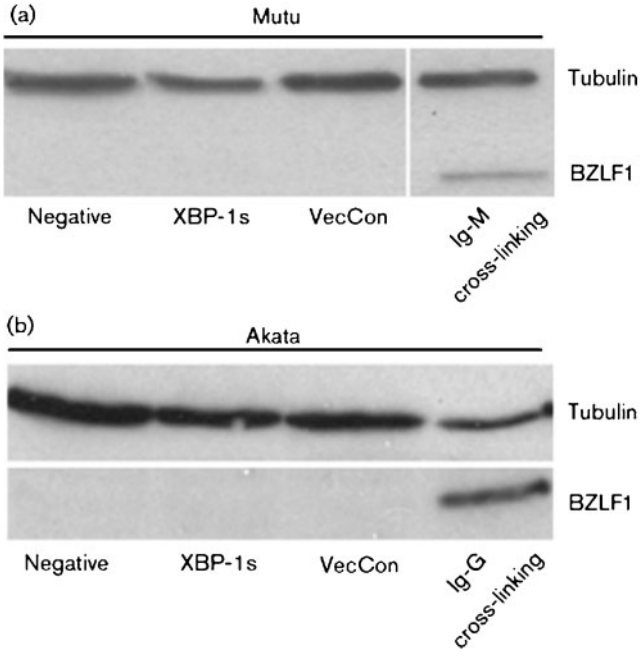


Fig. 2. Burkitt's lymphoma (BL) cell lines do not express BZLF1 after transduction with XBP-1s. (a) Mutu and (b) Akata, both EBV-positive BL cell lines, were transduced with XBP-1s and vector control (VecCon) but BZLF1 expression was not detected. B-cell surface cross-linking (BCR) with human Ig is used as positive control to induce BZLF1 expression: (a) Mutu, with Ig-M. (b) Akata, with Ig-G.

Overexpression of XBP-1s in BL or PEL cells does not induce EBV BRLF1 and BMRF1 expression

BRLF1 is a second EBV immediate–early protein whose expression is induced by BZLF1 during the EBV lytic cycle. Together these two proteins induce the rest of EBV lytic cascade. We examined the expression of BRLF1 after overexpressing XBP-1s in both BL (Akata) (Fig. 3a) and PEL (JSC-1) (Fig. 3b) cells. The expression of the EBV early protein BMRF1, also known as diffused early antigen (EA-D) (Bayliss & Wolf, 1981), was also examined by using Western blot analysis. BMRF1 transcription is activated by BZLF1 (Kenney *et al.*, 1992). However, in BL or PEL cell lines, overexpression of XBP-1s did not induce BZLF1, BRLF1 and BMRF1 expression. Interestingly, co-induction of K-RTA, BZLF1 and BRLF1 by TPA in PEL failed to induce BMRF1 expression, consistent with mutual co-repression of the full virus lytic cycle (Jiang *et al.*, 2008).

EBV lytic-cycle promoters do not respond to XBP-1s

Previously, we showed that XBP-1s can transactivate the K-RTA promoter (Wilson *et al.*, 2007), thereby inducing the KSHV lytic cycle. We therefore determined whether the

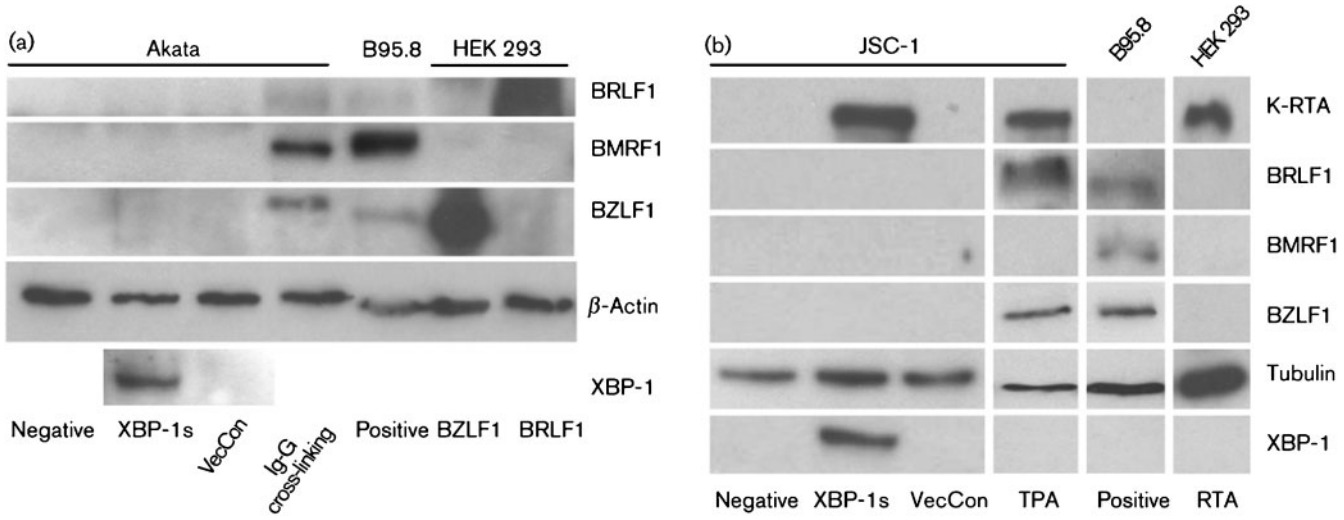


Fig. 3. Expression of XBP-1s does not induce BRLF1 and BRLF1 protein expression in either BL or PEL cell lines. (a) Akata and (b) JSC-1 cell lines were transduced with XBP-1s and vector control with m.o.i. of 5 and 2, respectively (m.o.i. measured in JSC-1 cells). Neither BZLF1, BRLF1 nor BMRF1 expression were detected in either cell line. BCR cross-linking was used as a positive control for Akata to induce EBV lytic gene expression and TPA was used as a positive control for JSC-1 cell lines to induce KSHV and EBV lytic gene expression. Both positive controls showed strong BZLF1 expression, weaker BRLF1 expression, and, for Akata (a) only, expression of BMRF1 with BCR cross-linking. B95.8 cell line was used as a positive control for EBV immediate–early and early protein expression.

EBV BZLF1 or BRLF1 promoters can be transactivated by XBP-1s. A construct containing BZLF1 promoter fused to a luciferase reporter gene (McDonald *et al.*, 2010) was used to determine the response to XBP-1s in HEK 293T cells; however, we were unable to show strong XBP-1s activity on the BZLF1 promoter (Fig. 4a). The experiments were repeated using the BRLF1 promoter (Bhende *et al.*, 2007) in either the methylated or unmethylated form. Both forms of the BRLF1 promoter were transactivated by BRLF1 and had a lesser but detectable response to XBP-1s (Fig. 4b). The activity was therefore independent of the promoter methylation status. The activity of XBP-1s in this assay system was confirmed by using a KSHV RTA luciferase promoter reporter that responded to both KSHV RTA and XBP-1s (Fig. 4c). The expression of XBP-1s in the luciferase assay was confirmed by Western blot analysis (Fig. 4a, b insert).

XBP-1 is unspliced in different B-cell lymphomas and splicing is not induced by TPA or valporic acid (VPA)

XBP-1u is expressed in both PEL and BL cell lines. DTT, a reducing agent able to induce the KSHV lytic cycle in KSHV-infected HEK 293T cells, can cause the splicing of inactive XBP-1u mRNA to active XBP-1s mRNA by promoting the unfolded-protein response (UPR) (Wilson *et al.*, 2007). As TPA can induce K-RTA and BZLF1 expression, and previous reports show other chemical agents can induce these viruses (Ye *et al.*, 2007), we determined whether they do so in part by inducing XBP-1splicing. DTT, TPA, VPA and sodium butyrate (NaB) (both histone deacetylase inhibitors) were used and XBP-1 splicing was assessed in the PEL cell line JSC-1 and BC3 cl6 cell line. Only DTT caused XBP-1 splicing in PEL cell lines despite TPA being capable of inducing KSHV K-RTA and EBV BZLF1 expression (Fig. 5b, c, f). We also examined the splicing status of XBP-1 in the BL cell lines Akata and Mutu. Again, XBP-1 is expressed as the unspliced inactive form and this only undergoes splicing in the presence of DTT (Fig. 5d, e).

Induction of XBP-1 splicing by the UPR does not induce the KSHV and EBV lytic cycle

As DTT is capable of inducing XBP-1 splicing and the KSHV lytic cycle in HEK 293T cells (Wilson *et al.*, 2007), we examined the ability of DTT to induce the lytic reactivation of KSHV and EBV by Western blotting for the expression of K-RTA and BZLF1. Despite being able to cause XBP-1 splicing, DTT resulted in large amounts of cell death in PEL (Fig. 5h) After treating the cells with DTT for 10 min, JSC-1 cells lose membrane integrity and stain with trypan blue; after 48 h of incubation, more than 50 % of JSC-1 cells are dead (Fig. 5h). Therefore, and because of the toxicity of DTT in PEL cells, it is not possible to access the effect of XBP-1 splicing (Fig. 5f). The BL cell line Akata was more resilient to DTT treatment; however, despite the ability of DTT to induce XBP-1 splicing, BZLF1 was not expressed (Fig. 5g).

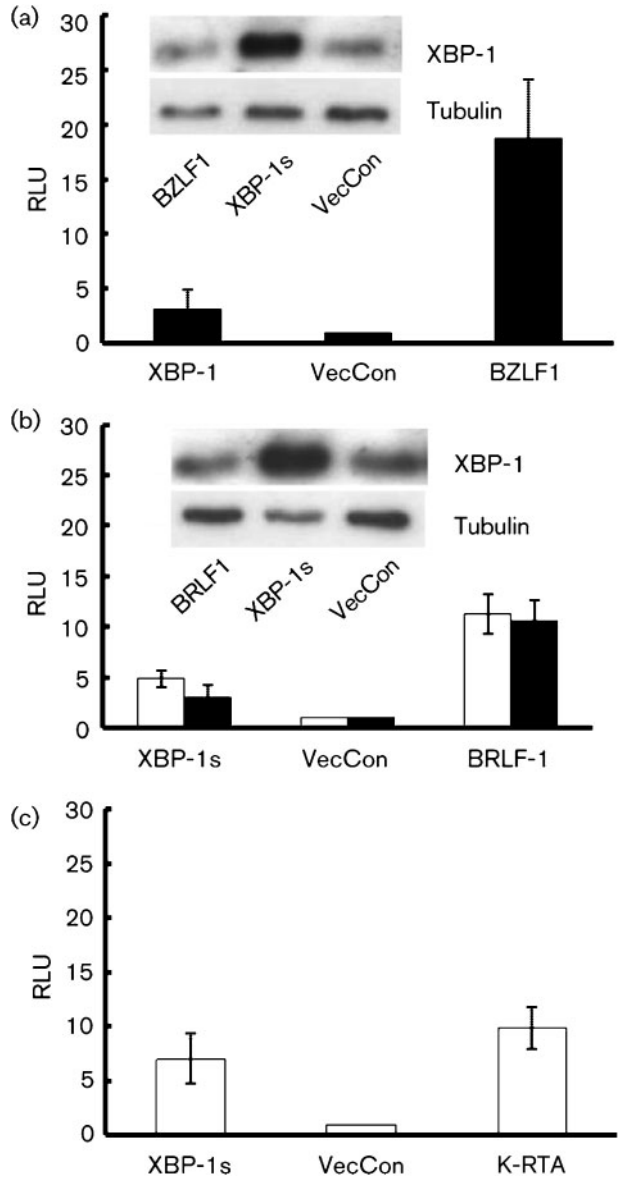


Fig. 4. BZLF1 and BRLF1 promoter responses to XBP-1s. (a) XBP-1s weakly transactivates the Zp in a luciferase-reporter assay in HEK 293T cells. BZLF1 overexpression transactivates the Zp as a positive control. (b) BRLF1 and XBP-1s expression transactivate the Rp. The XBP-1s expression level in Zp- (d, insert) and Rp- (e, insert) transfected cells co-transfected with BZLF1, BRLF1 or XBP-1s was determined by Western blotting. (c) The K-RTA promoter was used as a positive control for XBP-1s expression and activity. The promoter is transactivated by both K-RTA and XBP-1s. RLU, Relative light units.

Overexpression of XBP-1s or use of chemical agents does not induce expression of BZLF1 and BRLF1 mRNA in PEL or BL

Although XBP-1s overexpression did not induce detectable BZLF1 and BRLF1 protein expression in PEL or BL cell lines, the luciferase assays showed that XBP-1s weakly transactivated Zp and Rp in HEK 293T cells. We therefore performed a quantitative real-time PCR (Q-RT-PCR) to determine the mRNA levels of BZLF1 and BRLF1 in PEL and BL cell lines in response to XBP-1s and various chemical inducers of the KSHV and EBV lytic cycles.

In Akata cells, BZLF1 and BRLF1 mRNA expression did not increase following XBP-1s expression or following TPA, VPA and DTT treatment (Fig. 6a). BCR cross-linking used as a positive control induced BZLF1 mRNA expression significantly ($P<0.05$, two-tailed *t*-test). The increase of BRLF1 mRNA is not statistically significant ($P>0.05$, two-tailed *t*-test) and is consistent with the lack of BRLF-1 protein expression (Fig. 3a). This is also consistent with previous work by Amon *et al.* (2004), showing that BCR-cross-linking transactivates Zp, but not Rp.

EBV BZLF1, EBV BRLF1 and KSHV K-RTA mRNA expression was determined in JSC-1 cells by using Q-RT-PCR (Fig. 6b). After XBP-1s transduction or VPA and DTT treatment, BZLF1 and BRLF1 mRNA expression was not significantly different from control cells ($P>0.05$, two-tailed *t*-test). However, the K-RTA mRNA level increased significantly in response to XBP-1s, TPA, VPA and DTT ($P<0.05$, two-tailed *t*-test). TPA induced a significant increase in the BZLF1 mRNA level, compared with the negative control ($P<0.05$, two-tailed *t*-test). These data confirm that XBP-1s alone does not transactivate the EBV Zp or Rp in PEL or BL cell lines.

XBP-1s and protein kinase D together do not induce BZLF1 expression in PEL cell lines

Previously Bhende *et al.* (2007) showed that XBP-1s alone was not sufficient to induce lytic reactivation of EBV and that PKD was also required. We therefore performed luciferase assays to investigate the effect of the combined expression of XBP-1s and a constitutively active PKD (pPKDm-IG). In HEK 293T cells, PKD alone weakly transactivated the BZLF1 promoter but not the BRLF1 promoter (Fig. 7a). Conversely, XBP-1s in combination with PKD weakly transactivated the BRLF1 promoter (Fig. 7a). In order to ensure that the lack of a robust effect from PKD is not cell type specific, we performed the luciferase assay in HeLa cells for the Zp. Here, XBP-1s and active PKD alone do not transactivate Zp, but together weakly transactivate Zp (Fig. 7b). However, these effects are not statistically significant ($P>0.05$, two-tailed *t*-test). At all times BZLF1 and BRLF1 were able to transactivate their respective promoters.

Importantly, overexpressing either PKD or XBP-1s alone or together in JSC-1 cells did not induce BZLF1 or BRLF1 protein expression (Fig. 7c), whereas K-RTA expression was induced whenever XBP-1s was overexpressed. Q-RT-PCR confirmed that the mRNA expression of BZLF1 and BRLF1 did not increase when PKD was expressed with and without XBP-1s (Fig. 7d). K-RTA mRNA expression increased significantly ($P<0.05$, two-tailed *t*-test) whenever XBP-1s

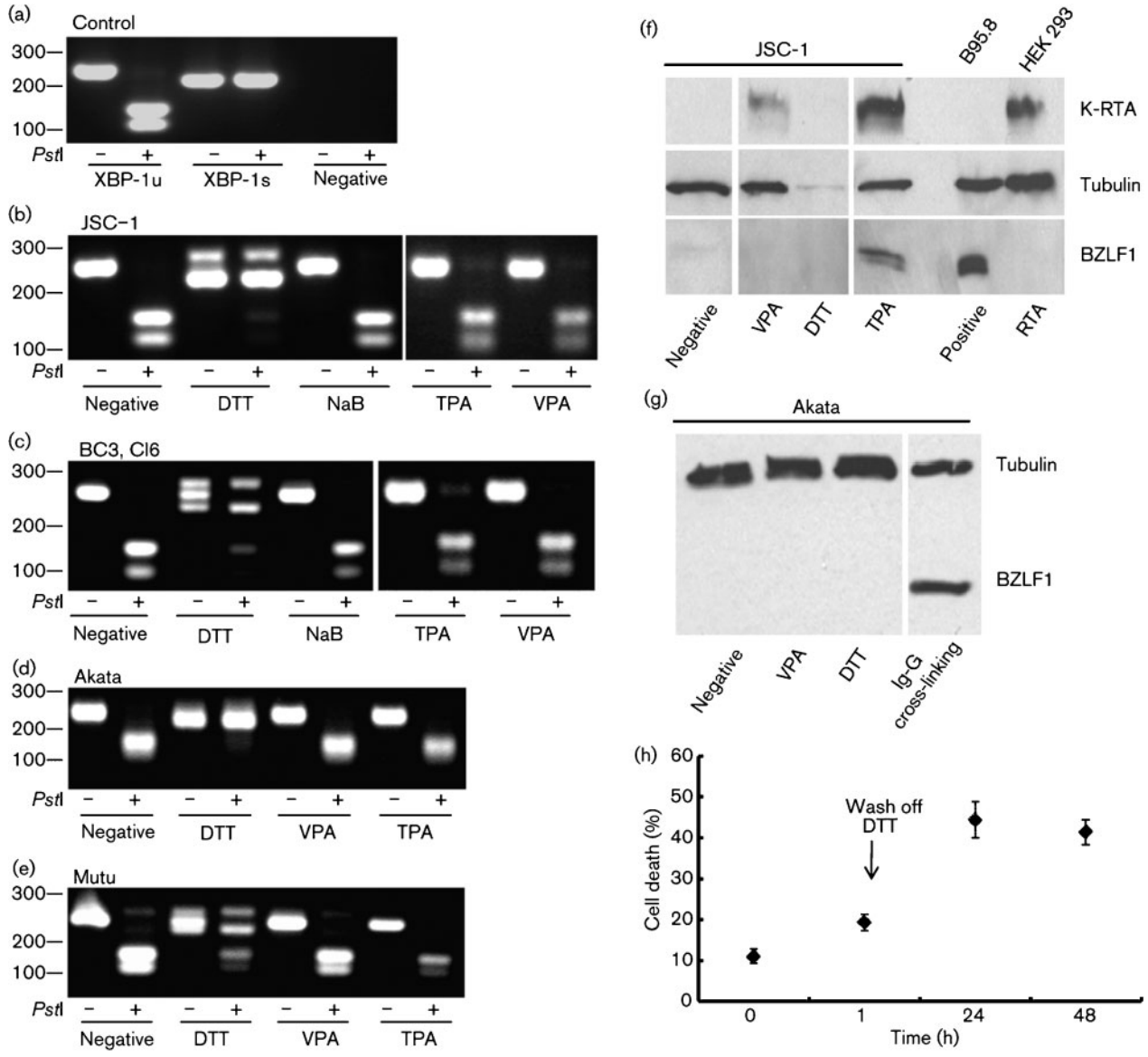


Fig. 5. NaB, TPA and VPA do not induce XBP-1 splicing in PEL and BL cell lines. (a) XBP-1 is expressed as an unspliced (XBP-1u) form, but the spliced, active form (XBP-1s) can be detected following ER stress induction by the resistance to *Pst*I digestion of the XBP-1s PCR product. Treating the cells with DTT induces ER stress and XBP-1u splicing in all lymphoma cell lines (b–e). However, NaB, TPA and VPA treatment do not induce XBP-1 splicing in the lymphoma cell lines (b–e) even though VPA induces K-RTA expression in the PEL JSC-1 and TPA induces both K-RTA and BZLF1 in JSC-1 cells (f). VPA and DTT do not induce BZLF1 expression in the BL line Akata (g). DTT toxicity (h) results in PEL cell death and the absence of detectable proteins (f).

was overexpressed (Fig. 7d). Therefore, active PKD alone does not lead to EBV BZLF1 and EBV BRLF1, or KSHV K-RTA expression. Also, active PKD together with XBP-1s does not induce EBV BZLF1 and BRLF1 expression.

DISCUSSION

Plasma cell terminal differentiation and the associated expression of XBP-1s are linked to EBV and KSHV lytic cycle induction (Bhende *et al.*, 2007; Laichalk & Thorley-Lawson, 2005; Sun & Thorley-Lawson, 2007; Wilson *et al.*, 2007; Yu *et al.*, 2007). Recently this has been extended beyond human gammaherpesviruses with the demonstration that plasma cells account for most of the lytic reactivation of murine herpesvirus 68 in mice (Liang *et al.*, 2009). In EBV and KSHV co-infected PEL however, selective induction of one viral lytic cycle cross-represses the other viral lytic cycle and co-expression of both KSHV

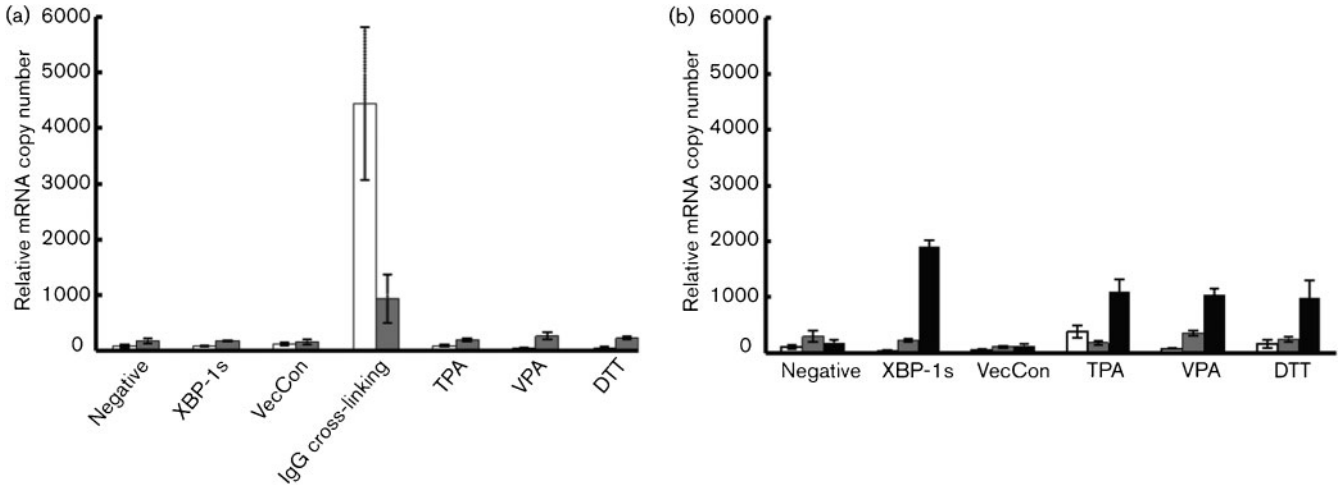


Fig. 6. Overexpression of XBP-1s in PEL and BL cell lines does not increase the mRNA expression levels of BZLF1 and BRLF1. Q-RT-PCR was used to determine the expression of BZLF1 mRNA (open bars), BRLF1 mRNA (grey bars) and K-RTA mRNA (black bars) after transduction with XBP-1s lentivirus or treatment with chemical inducers for (a) Akata and (b) JSC-1 cell lines. (a) Neither XBP-1s overexpression nor TPA, VPA or DTT treatments were able to increase BZLF1 or BRLF1 mRNA expression. BCR cross-linking induced BZLF mRNA expression ($P<0.05$, two-tailed *t*-test). (b) In JSC-1 cells, expression of XBP-1s only increased K-RTA mRNA expression. TPA, VPA and DTT treatment also increased K-RTA mRNA expression significantly ($P<0.05$, two tailed *t*-test). TPA-treated JSC-1 cells expressed BZLF1 mRNA compared with negative control ($P<0.05$, two tailed *t*-test).

and EBV immediate–early proteins results in mutual inhibition (Jiang *et al.*, 2008; Miller *et al.*, 1997; Xu *et al.*, 2007). This suggests that a selective lytic switch that induces one or other virus is required for successful EBV or KSHV lytic replication in PEL. Here we show that XBP-1s is able to induce KSHV lytic replication alone in dually infected PEL. XBP-1s does not induce the expression of EBV BZLF1 and BRFL1 immediate–early–proteins or transcripts in PEL or BL. XBP-1s with PKD also does not induce the expression of BZLF1 and BRFL1 immediate–early proteins or transcripts in PEL.

Previous studies have suggested XBP1s, either alone or in combination with PKD, can activate the EBV lytic cycle (Bhende *et al.*, 2007; Sun & Thorley-Lawson, 2007); however, the mechanistic detail is unclear. Sun & Thorley-Lawson (2007) showed that, in HeLa cells, XBP-1s cannot transactivate Zp. Whereas, Bhende *et al.* (2007) showed, also in HeLa cells, that XBP-1s weakly transactivates Zp and Rp, and the effect on Zp could be enhanced by co-expression of constitutively active PKD (Bhende *et al.*, 2007; Sun & Thorley-Lawson, 2007). Our data in HEK 293T cells supports the observations that XBP-1s alone can weakly transactivate Zp and Rp. Similarly, we show co-expression of XBP-1s and PKD weakly transactivates Zp in HEK 293T and HeLa cells, although not to the magnitude described previously. This may reflect differences in assay sensitivity or may reflect the fact that we have used human XBP-1s in our studies rather than the murine XBP-1s used by Bhende *et al.* (2007). Nevertheless, the relevance of these observations in HEK 293T and HeLa cells, and the response of EBV to XBP-1s and PKD in B-cell tumour lines, is questionable.

The ability of XBP-1s to transactivate either Zp or Rp probably depends on both the cell type and on the nature of the individual cell lines. This is supported by observations that distinct chemical inducing agents have different effects on the induction of EBV and KSHV lytic cycles in various lymphoma lines (Countryman *et al.*, 2008; Miller *et al.*, 1997). In a lymphoblastoid and a nasopharyngeal carcinoma cell line XBP-1s and PKD clearly activated EBV BZLF1 protein expression, but BL or PEL cell lines were not tested (Bhende *et al.*, 2007). In an unusual multiple-myeloma cell line latently infected with EBV, only low levels of BZLF1 transcript could be induced by XBP-1s. This observation contrasts with recent data indicating that EBV can latently infect multiple myeloma cell lines *in vitro*. In these circumstances the endogenous, active XBP-1s does not drive the EBV lytic cycle (Anastasiadou *et al.*, 2009). Here we show that despite transactivation of Zp and Rp by XBP-1s in HEK 293T cells, there are no detectable levels of BZLF1 and BRLF1 mRNA or proteins after XBP-1s overexpression in B-cell tumour lines.

Why EBV Zp is refractory to XBP-1s in PEL and BL, even when PKD is co-expressed, is not clear, although the role of the chromatin structure at Zp is likely to be important. Indeed, the binding of myocyte enhancer factor 2D to Zp recruits class II histone deacetylases (HDACs), which presumably promote a repressed chromatin structure on Zp in lymphoma cell lines (Gruffat *et al.*, 2002; McDonald

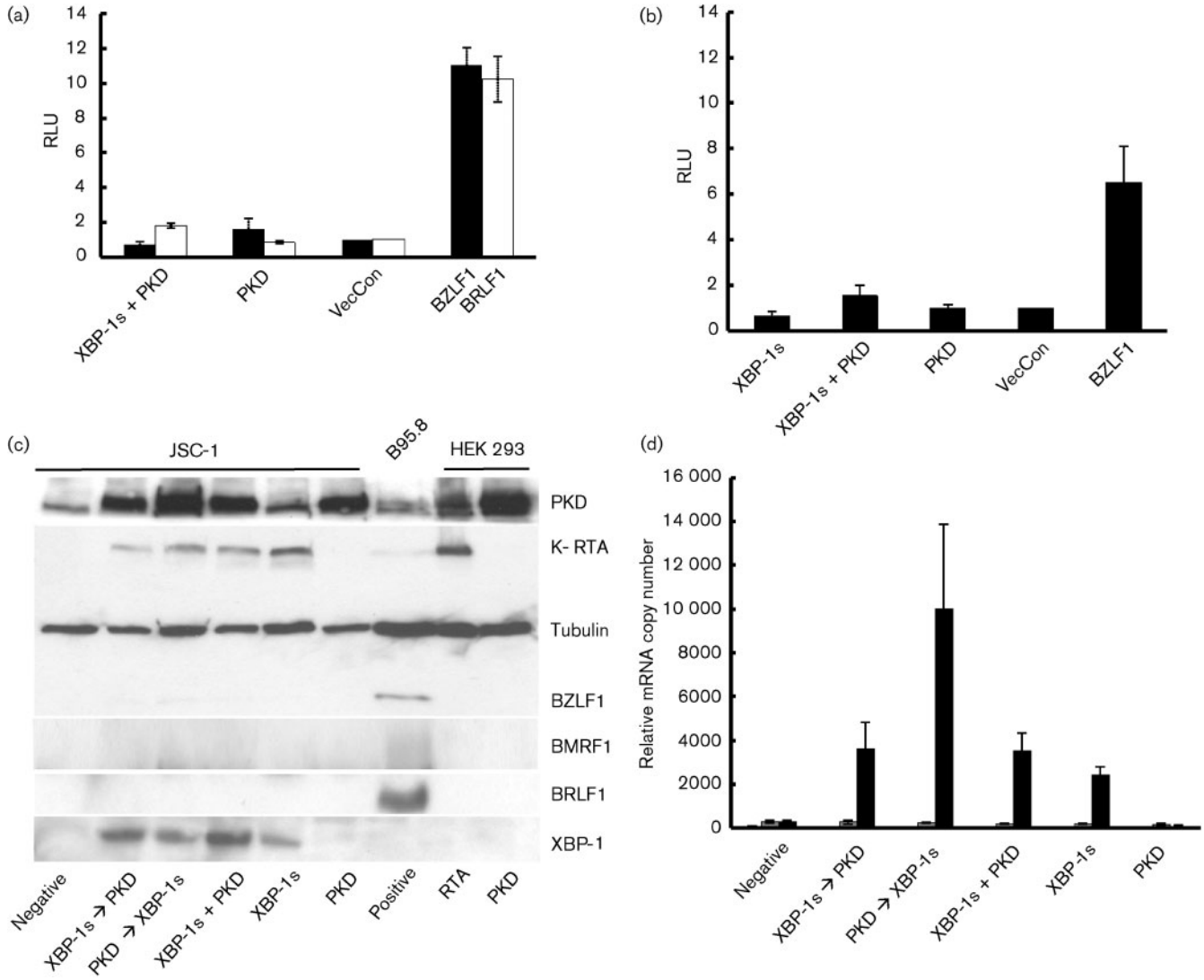


Fig. 7. Co-expression of XBP-1s and active PKD does not transactivate Zp or induce BZLF1 expression. (a) HEK 293-T-cells were transduced with Zp (black bars) or Rp (open bars) luciferase-reporter vectors with either active PKD or XBP-1s together with active PKD. Active PKD alone shows weak Zp activity and together with XBP-1s weak Rp activity compared to control empty vector (VecCon) and BZLF1 or BRLF1 expression vectors. However, the increase of luciferase activity was not statistically significant compared to vector control ($P>0.05$, two-tailed *t*-test). (b) The Zp promoter assays were repeated in HeLa cells where only XBP-1s and PKD together showed weak promoter activity compared with BZLF1 overexpression. However, this was not statistically significant compared to vector control ($P>0.05$, two tailed *t*-test). RLU, Relative light units. (c) Transduction of JSC-1 cells with XBP-1s and active PKD either alone, in combination or sequentially (denoted by arrows) does not induce BZLF1, BRLF1 or BMRF1 expression in PEL, but all combinations that include XBP-1s induce K-RTA expression. (d) mRNA expression in JSC-1, measured by Q-RT-PCR after XBP-1s and/or active PKD transduction for BZLF1 mRNA (open bars), BRLF1 mRNA (grey bars) and K-RTA (black bars). Only K-RTA mRNA expression increased significantly ($P<0.05$, two tailed *t*-test) after XBP-1s transduction with or without PKD expression.

et al., 2010). Derepression of Zp-associated chromatin by PKD was suggested as the mechanism that allows XBP-1s to activate Zp (Bhende *et al.*, 2007). PKD is a member of the serine/threonine protein kinase family (Rey *et al.*, 2006) which can be activated by protein kinase C (PKC) (Zugaza *et al.*, 1996), a cellular target of phorbol esters such as TPA (Wang, 2006). PKD can also be activated by BCR cross-linking via the PKC pathway (Matthews *et al.*, 2000). In our hands, TPA induces both K-RTA and BZLF1 in PEL and BCR cross-linking induces BZLF1 in BL. Activated PKD can phosphorylate and therefore inactivate class IIa HDACs; this is similar to the effect of HDAC inhibitors (HDACi), such as trichostatin A (Chang & Liu, 2000), VPA (Feng & Kenney, 2006) and sodium butyrate (NaB). As

HDACi are also capable of inducing EBV lytic reactivation in various cell lines (Gradoville *et al.*, 2002), we investigated whether XBP-1s is also induced by any of these chemical inducers. Here we show that neither NaB, VPA nor TPA induce activated XBP-1s in PEL, and therefore the effects of HDACi and TPA on the Zp is independent of XBP-1s.

BCR cross-linking induces the EBV lytic cycle and can also transiently induce XBP-1s production. However, this transient XBP-1s cannot be unambiguously linked to BZLF1 expression (McDonald *et al.*, 2010). In this study we used DTT to chemically induce XBP-1s in PEL and BL cell lines. Although DTT is able to induce XBP-1 splicing and K-RTA in an XBP-1s dependent manner in HEK 293T cells (Wilson *et al.*, 2007), it causes a large amount of cell death in PEL cells, making it impossible to determine K-RTA, BZLF1 or BRFL1 protein expression. Nevertheless, K-RTA mRNA expression was detected following DTT treatment of PEL (Fig. 6b). Also DTT does not cause cell death in BL cells; it induces XBP-1s production but does not induce BZLF1 expression. Taken together these data suggest that induction of the EBV lytic cycle in lymphoma cell lines of PEL and BL is not linked to XBP-1s activity, although the effects in B-cells *in vivo* may be different.

METHODS

Cell culture. The PEL cell line JSC-1 and the BL cell lines Mutu, Daudi and Akata were grown in RPMI 1640 medium (Invitrogen) with 10 % FCS (BioSera), 100 units penicillin ml^{-1} and 100 units streptomycin ml^{-1} (Invitrogen) at 37 °C in 5 % CO_2. All super-infected PEL cell lines, CRO6 clone 2, BC3 clone 6 and BC3 clone 10 (a kind gift from Pankaj Treviti), were grown with G418 selection as described previously (Xu *et al.*, 2007). HEK 293T cells and HeLa cells were grown in Dulbecco's modified Eagle's medium (Invitrogen) with 15 and 10 % FCS, respectively, 100 units penicillin ml^{-1} and 100 units streptomycin ml^{-1} (Invitrogen) at 37 °C in 10 % CO_2. To induce BZLF1 and K-RTA expression, cells were cultured with the following concentrations of inducing agents: DTT, 2 mM; goat anti-human IgG antibody (Sigma), 5 μg ml^{-1}; 12-*O*-tetradecanoyl-phorbol 13-acetate (TPA; Sigma), 20 ng μl^{-1}; VPA (Merck), 1.5 mM.

Lentiviral vector construction and transduction. The lentiviral vector construction of pXBP1sIG and pIG has been described previously (Wilson *et al.*, 2007). The lentiviral vector for activated PKD, pPKDm-IG, was produced from the pHA.PKD.S738A/S742A plasmid kindly supplied by Alex Toker (Storz & Toker, 2003) (Addgene). Briefly, *Bam*HI and *Xho*I were used to clone the HA.PKD.S738A/S742A insert into the lentiviral vector pIG to produce the pPKDm–IG construct. To produce lentiviral viruses, HEK 293T cells were transfected with 1 μg pMDG, 1 μg p8.91 and 1.5 μg of lentiviral vector, by using FuGENE-6 (Roche). The supernatants were collected 48 and 72 h post-transfection and filtered. Both PEL and BL cell lines (5×10^4 cells per well in a 24-well plate) were transduced with lentivirus vector at an input equivalent to an m.o.i. of 2 or 5 as measured in JSC-1 cells. This resulted in an actual level of infection of 60–80 % in PEL and 10–25 % in BL cell lines. Lentiviruses were added to the cells, spinoculated for 1 h, at 500 ***g*** and room temperature. No selection for infected cells was used after the transduction. Forty-eight hours after transduction the cells were analysed using flow cytometry.

RNA extraction and reverse-transcriptase PCR (RT-PCR). Total RNA was purified from $8–10 \times 10^5$ cells resuspended in TRIzol (Invitrogen). The TRIzol mixture was first treated with chloroform and RNA was isolated using an RNA extraction kit (Qiagen), including an on-column DNase (Promega) digestion. Reverse transcription was carried out using Ominiscript Reverse Transcriptase (Qiagen) according to the manufacturer's instructions with 1–2 μg total RNA.

PCR and restriction digestion. Oligo-dT (Promega)-primed cDNA was used for PCR amplification across the XBP-1 intron as described previously by Wilson *et al.* (2007). The PCR product was then digested with *Pst*I for 1 h at 37 °C.

Q-RT-PCR for mRNA. Q-RT-PCR was performed with a QuantiTect SYBR Green PCR kit (Qiagen) using the following primers: BZLF1 (5′-CTATCAGGACCTGGGAGGGC-3′ and 5′-CACAGCACACAA-GGCAAAGG-3′) (Schelcher *et al.*, 2005), BRLF1 (5′-AATTTAC-AGCCGGGAGTGTG-3′ and 5′-AGCCCGTCTTCTTACCCTGT-3′) (Chia *et al.*, 2008), K-RTA (5′-TTGGTGCGCTATGTGGTCTG-3′ and 5′-GGAAGGTAGACCGGTTGGAA-3′) (Caselli *et al.*, 2005) and L32 (5′-CAACATTGGTTATGCAAGCAACA-3′ and 5′-TGACGTTG-TGGACCAGGAACT-3′) (Schelcher *et al.*, 2005). L32 is a cellular ribosomal gene used to assess the preparation of cDNA and all the mRNA levels were normalized with L32. The PCR was prepared according to the manufacturer's instructions and was performed by using an ABI Prism 7000 (Applied Biosystems) with the following programme: 50 °C for 2 min, 95 °C for 15 min, 95 °C for 15 s and 60 °C for 1 min; the third and fourth steps are then repeated for 40 cycles.

Luciferase promoter assays. HEK 293T cells or HeLa cells were plated at a density of 2×10^4 cells per well in 96-well plates. Cells were transfected the next day with 20 ng of the appropriate reporter and expression constructs by using fuGENE-6 (Roche). The ORF50 reporter plasmids were a kind gift from Erle Robertson (Robertson & Ambinder, 1997). The BZLF1 promoter and expression vectors have been described previously (Bryant & Farrell, 2002) and the BRLF1 promoter and expression constructs were a kind gift from Shannon Kenney (Bhende *et al.*, 2007). The pXBP1sIG and pIG plasmids were described previously (Wilson *et al.*, 2007). Co-transfection of 2 ng of Renilla expression vector (a kind gift from Professor Gary Stein, University of Massachusetts, MA, USA) was also used for monitoring transfection efficiency. After 48 h, the relative light units (RLU) and the expression of the Renilla construct were determined using luciferase Stop & Glo reagents (dual-luciferase assay kit; Promega) according to the manufacturer's instructions. Data were read using a GloMaz96 microplate luminometer (Promega) with a single injector. All assays were performed in triplicate.

Methylation of promoter luciferase constructs. BRLF1 promoter was methylated using CpG methyltransferase (New England Biolabs) according to the manufacturer's instructions. Methylation of the plasmids was confirmed by performing restriction enzyme digestion with methylation-sensitive enzymes. Rp was digested with *BgI*I and *Mln*I.

Immunoblotting. Forty-eight hours after transduction (with no selection), the samples were lysed with a mixture of 1 M DTT and sample buffer (1 : 3 ratio) (0.2 M Tris-HCL pH 6.8, 5.2 % SDS, 20 % Glycerol and bromophenol blue) then sonicated and heated to 95 °C for 5 min. Proteins were resolved on 10 % polyacrylamide gels by SDS-PAGE and transferred to a PVDF membrane. The membrane was blocked with 5 % non-fat dried milk powder (NFDM) in TBS solution containing 0.1 % Tween 20 (Sigma) (TBS-T) for 1 h. The membrane was then probed with primary antibody as follows: BZLF1, BZ-1 diluted 1 : 2000 (Bryant & Farrell, 2002); tubulin, anti-α-tubulin

clone DM1A (Sigma) diluted 1 : 40,000; K-RTA polyclonal, kind gift of Don Ganem (Lukac *et al.*, 1998), diluted 1 : 40 000; PKD, PKC μ(D-20), sc-935 (Santa Cruz) diluted 1 : 2000; BRLF1, anti-EBV transcription factor R (Argene) diluted 1 : 2000; BMRF1, EBV early antigen diffuse (Vector Laboratories) diluted 1 : 2000; XBP-1s, a kind gift from Giovanna Roncador, diluted 1 : 50 (Maestre *et al.*, 2009), all in 1 % NFDM in TBS-T solution at 4 °C overnight. The membrane was then incubated with appropriate HRP-conjugated secondary antibodies (GE Healthcare) in 1 % NFDM TBS-T solution for 1 h at room temperature. The blots were washed five times for 5 min periods in TBS-T before developing, and were visualized using ECL Western blotting detection reagents or ECL Advanced Western blotting detection reagents (GE Healthcare).

ACKNOWLEDGEMENTS

We thank Pankaj Treviti, Erle Robertson, Shannon Kenney, Alison Sinclair, Giovanna Roncador and Don Ganem for providing reagents for this work; the VGB group, especially Ed Tsao at the Windeyer building at UCL for discussion, technical support and advice. I. Y.-C. Lai is supported by Cancer Research UK and Paul Kellam is supported by grants from Cancer Research UK and the Wellcome Trust.

REFERENCES

Amon, W., Binne, U. K., Bryant, H., Jenkins, P. J., Karstegl, C. E. & Farrell, P. J. (2004). Lytic cycle gene regulation of Epstein-Barr virus. *J Virol* **78**, 13460–13469.

Anastasiadou, E., Vaeth, S., Cuomo, L., Boccellato, F., Vincenti, S., Cirone, M., Presutti, C., Junker, S., Winberg, G. & other authors (2009). Epstein–Barr virus infection leads to partial phenotypic reversion of terminally differentiated malignant B cells. *Cancer Lett* **284**, 165–174.

Babcock, G. J., Decker, L. L., Volk, M. & Thorley-Lawson, D. A. (1998). EBV persistence in memory B cells *in vivo*. *Immunity* **9**, 395–404.

Babcock, G. J., Decker, L. L., Freeman, R. B. & Thorley-Lawson, D. A. (1999). Epstein–Barr virus-infected resting memory B cells, not proliferating lymphoblasts, accumulate in the peripheral blood of immunosuppressed patients. *J Exp Med* **190**, 567–576.

Bayliss, G. J. & Wolf, H. (1981). The regulated expression of Epstein–Barr virus. III. Proteins specified by EBV during the lytic cycle. *J Gen Virol* **56**, 105–118.

Bhende, P. M., Dickerson, S. J., Sun, X., Feng, W. H. & Kenney, S. C. (2007). X-box-binding protein 1 activates lytic Epstein-Barr virus gene expression in combination with protein kinase D. *J Virol* **81**, 7363–7370.

Bryant, H. & Farrell, P. J. (2002). Signal transduction and transcription factor modification during reactivation of Epstein-Barr virus from latency. *J Virol* **76**, 10290–10298.

Calfon, M., Zeng, H., Urano, F., Till, J. H., Hubbard, S. R., Harding, H. P., Clark, S. G. & Ron, D. (2002). IRE1 couples endoplasmic reticulum load to secretory capacity by processing the XBP-1 mRNA. *Nature* **415**, 92–96.

Caselli, E., Galvan, M., Cassai, E., Caruso, A., Sighinolfi, L. & Di Luca, D. (2005). Human herpesvirus 8 enhances human immunodeficiency virus replication in acutely infected cells and induces reactivation in latently infected cells. *Blood* **106**, 2790–2797.

Chang, L. K. & Liu, S. T. (2000). Activation of the BRLF1 promoter and lytic cycle of Epstein–Barr virus by histone acetylation. *Nucleic Acids Res* **28**, 3918–3925.

Chia, M. C., Leung, A., Krushel, T., Alajez, N. M., Lo, K. W., Busson, P., Klamut, H. J., Bastianutto, C. & Liu, F. F. (2008). Nuclear factor-Y and Epstein–Barr virus in nasopharyngeal cancer. *Clin Cancer Res* **14**, 984–994.

Countryman, J. & Miller, G. (1985). Activation of expression of latent Epstein–Barr herpesvirus after gene transfer with a small cloned subfragment of heterogeneous viral DNA. *Proc Natl Acad Sci U S A* **82**, 4085–4089.

Countryman, J. K., Gradoville, L. & Miller, G. (2008). Histone hyperacetylation occurs on promoters of lytic cycle regulatory genes in Epstein-Barr virus-infected cell lines which are refractory to disruption of latency by histone deacetylase inhibitors. *J Virol* **82**, 4706–4719.

Dalton-Griffin, L., Wilson, S. J. & Kellam, P. (2009). X-box binding protein 1 contributes to induction of the Kaposi's sarcoma-associated herpesvirus lytic cycle under hypoxic conditions. *J Virol* **83**, 7202–7209.

Fan, W., Bubman, D., Chadburn, A., Harrington, W. J., Jr, Cesarman, E. & Knowles, D. M. (2005). Distinct subsets of primary effusion lymphoma can be identified based on their cellular gene expression profile and viral association. *J Virol* **79**, 1244–1251.

Feng, W. H. & Kenney, S. C. (2006). Valproic acid enhances the efficacy of chemotherapy in EBV-positive tumors by increasing lytic viral gene expression. *Cancer Res* **66**, 8762–8769.

Gires, O., Zimber-Strobl, U., Gonnella, R., Ueffing, M., Marschall, G., Zeidler, R., Pich, D. & Hammerschmidt, W. (1997). Latent membrane protein 1 of Epstein-Barr virus mimics a constitutively active receptor molecule. *EMBO J* **16**, 6131–6140.

Gradoville, L., Kwa, D., El-Guindy, A. & Miller, G. (2002). Protein kinase C-independent activation of the Epstein-Barr virus lytic cycle. *J Virol* **76**, 5612–5626.

Gruffat, H., Manet, E. & Sergeant, A. (2002). MEF2-mediated recruitment of class II HDAC at the EBV immediate early gene BZLF1 links latency and chromatin remodeling. *EMBO Rep* **3**, 141–146.

Jiang, Y., Xu, D., Zhao, Y. & Zhang, L. (2008). Mutual inhibition between Kaposi's sarcoma-associated herpesvirus and Epstein-Barr virus lytic replication initiators in dually-infected primary effusion lymphoma. *PLoS One* **3**, e1569.

Joseph, A. M., Babcock, G. J. & Thorley-Lawson, D. A. (2000). Cells expressing the Epstein-Barr virus growth program are present in and restricted to the naive B-cell subset of healthy tonsils. *J Virol* **74**, 9964–9971.

Kenney, S. C., Holley-Guthrie, E., Quinlivan, E. B., Gutsch, D., Zhang, Q., Bender, T., Giot, J. F. & Sergeant, A. (1992). The cellular oncogene c-myb can interact synergistically with the Epstein–Barr virus BZLF1 transactivator in lymphoid cells. *Mol Cell Biol* **12**, 136–146.

Laichalk, L. L. & Thorley-Lawson, D. A. (2005). Terminal differentiation into plasma cells initiates the replicative cycle of Epstein-Barr virus *in vivo*. *J Virol* **79**, 1296–1307.

Liang, X., Collins, C. M., Mendel, J. B., Iwakoshi, N. N. & Speck, S. H. (2009). Gammaherpesvirus-driven plasma cell differentiation regulates virus reactivation from latently infected B lymphocytes. *PLoS Pathog* **5**, e1000677.

Lukac, D. M., Renne, R., Kirshner, J. R. & Ganem, D. (1998). Reactivation of Kaposi's sarcoma-associated herpesvirus infection from latency by expression of the ORF 50 transactivator, a homolog of the EBV R protein. *Virology* **252**, 304–312.

Maestre, L., Tooze, R., Canamero, M., Montes-Moreno, S., Ramos, R., Doody, G., Boll, M., Barrans, S., Baena, S. & other authors (2009). Expression pattern of XBP1(S) in human B-cell lymphomas. *Haematologica* **94**, 419–422.

Matthews, S. A., Rozengurt, E. & Cantrell, D. (2000). Protein kinase D. A selective target for antigen receptors and a downstream target for protein kinase C in lymphocytes. *J Exp Med* **191**, 2075–2082.

McDonald, C., Karstegl, C. E., Kellam, P. & Farrell, P. J. (2010). Regulation of the Epstein–Barr virus Zp promoter in B lymphocytes during reactivation from latency. *J Gen Virol* **91**, 622–629.

Miller, G., Heston, L., Grogan, E., Gradoville, L., Rigsby, M., Sun, R., Shedd, D., Kushnaryov, V. M., Grossberg, S. & Chang, Y. (1997). Selective switch between latency and lytic replication of Kaposi's sarcoma herpesvirus and Epstein-Barr virus in dually infected body cavity lymphoma cells. *J Virol* **71**, 314–324.

Panagopoulos, D., Victoratos, P., Alexiou, M., Kollias, G. & Mosialos, G. (2004). Comparative analysis of signal transduction by CD40 and the Epstein-Barr virus oncoprotein LMP1 *in vivo*. *J Virol* **78**, 13253–13261.

Reimold, A. M., Iwakoshi, N. N., Manis, J., Vallabhajosyula, P., Szomolanyi-Tsuda, E., Gravallese, E. M., Friend, D., Grusby, M. J., Alt, F. & Glimcher, L. H. (2001). Plasma cell differentiation requires the transcription factor XBP-1. *Nature* **412**, 300–307.

Rey, O., Papazyan, R., Waldron, R. T., Young, S. H., Lippincott-Schwartz, J., Jacamo, R. & Rozengurt, E. (2006). The nuclear import of protein kinase D3 requires its catalytic activity. *J Biol Chem* **281**, 5149–5157.

Robertson, K. D. & Ambinder, R. F. (1997). Methylation of the Epstein-Barr virus genome in normal lymphocytes. *Blood* **90**, 4480–4484.

Rooney, C. M., Rowe, D. T., Ragot, T. & Farrell, P. J. (1989). The spliced BZLF1 gene of Epstein-Barr virus (EBV) transactivates an early EBV promoter and induces the virus productive cycle. *J Virol* **63**, 3109–3116.

Schelcher, C., Valencia, S., Delecluse, H. J., Hicks, M. & Sinclair, A. J. (2005). Mutation of a single amino acid residue in the basic region of the Epstein-Barr virus (EBV) lytic cycle switch protein Zta (BZLF1) prevents reactivation of EBV from latency. *J Virol* **79**, 13822–13828.

Souza, T. A., Stollar, B. D., Sullivan, J. L., Luzuriaga, K. & Thorley-Lawson, D. A. (2005). Peripheral B cells latently infected with Epstein–Barr virus display molecular hallmarks of classical antigen-selected memory B cells. *Proc Natl Acad Sci U S A* **102**, 18093–18098.

Storz, P. & Toker, A. (2003). Protein kinase D mediates a stress-induced NF-κB activation and survival pathway. *EMBO J* **22**, 109–120.

Sun, C. C. & Thorley-Lawson, D. A. (2007). Plasma cell-specific transcription factor XBP-1s binds to and transactivates the Epstein-Barr virus BZLF1 promoter. *J Virol* **81**, 13566–13577.

Thorley-Lawson, D. A. (2001). Epstein-Barr virus: exploiting the immune system. *Nat Rev Immunol* **1**, 75–82.

Thorley-Lawson, D. A. & Gross, A. (2004). Persistence of the Epstein–Barr virus and the origins of associated lymphomas. *N Engl J Med* **350**, 1328–1337.

Trivedi, P., Takazawa, K., Zompetta, C., Cuomo, L., Anastasiadou, E., Carbone, A., Uccini, S., Belardelli, F., Takada, K. & other authors (2004). Infection of HHV-8+ primary effusion lymphoma cells with a recombinant Epstein-Barr virus leads to restricted EBV latency, altered phenotype, and increased tumorigenicity without affecting TCL1 expression. *Blood* **103**, 313–316.

Wang, Q. J. (2006). PKD at the crossroads of DAG and PKC signaling. *Trends Pharmacol Sci* **27**, 317–323.

Wilson, S. J., Tsao, E. H., Webb, B. L., Ye, H., Dalton-Griffin, L., Tsantoulas, C., Gale, C. V., Du, M. Q., Whitehouse, A. & Kellam, P. (2007). X box binding protein XBP-1s transactivates the Kaposi's sarcoma-associated herpesvirus (KSHV) ORF50 promoter, linking plasma cell differentiation to KSHV reactivation from latency. *J Virol* **81**, 13578–13586.

Xu, D., Coleman, T., Zhang, J., Fagot, A., Kotalik, C., Zhao, L., Trivedi, P., Jones, C. & Zhang, L. (2007). Epstein-Barr virus inhibits Kaposi's sarcoma-associated herpesvirus lytic replication in primary effusion lymphomas. *J Virol* **81**, 6068–6078.

Ye, F. C., Blackbourn, D. J., Mengel, M., Xie, J. P., Qian, L. W., Greene, W., Yeh, I. T., Graham, D. & Gao, S. J. (2007). Kaposi's sarcoma-associated herpesvirus promotes angiogenesis by inducing angiopoietin-2 expression via AP-1 and Ets1. *J Virol* **81**, 3980–3991.

Yu, F., Feng, J., Harada, J. N., Chanda, S. K., Kenney, S. C. & Sun, R. (2007). B cell terminal differentiation factor XBP-1 induces reactivation of Kaposi's sarcoma-associated herpesvirus. *FEBS Lett* **581**, 3485–3488.

Yuan, J., Cahir-McFarland, E., Zhao, B. & Kieff, E. (2006). Virus and cell RNAs expressed during Epstein-Barr virus replication. *J Virol* **80**, 2548–2565.

Zugaza, J. L., Sinnett-Smith, J., Van Lint, J. & Rozengurt, E. (1996). Protein kinase D (PKD) activation in intact cells through a protein kinase C-dependent signal transduction pathway. *EMBO J* **15**, 6220–6230.

OPEN ACCESS Freely available online

PLoS one

Specific Capture and Whole-Genome Sequencing of Viruses from Clinical Samples

Daniel P. Depledge[1]*, Anne L. Palser[2], Simon J. Watson[2], Imogen Yi-Chun Lai[1,2], Eleanor R. Gray[1], Paul Grant[3], Ravinder K. Kanda[1], Emily Leproust[4], Paul Kellam[1,2], Judith Breuer[1]

1 Division of Infection and Immunity, University College London, London, United Kingdom, **2** Pathogen Genetics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, United Kingdom, **3** Department of Virology, University College London Hospitals NHS Trust, London, United Kingdom, **4** Agilent Technologies, Santa Rosa, California, United States of America

Abstract

Whole genome sequencing of viruses directly from clinical samples is integral for understanding the genetics of host-virus interactions. Here, we report the use of sample sparing target enrichment (by hybridisation) for viral nucleic acid separation and deep-sequencing of herpesvirus genomes directly from a range of clinical samples including saliva, blood, virus vesicles, cerebrospinal fluid, and tumour cell lines. We demonstrate the effectiveness of the method by deep-sequencing 13 highly cell-associated human herpesvirus genomes and generating full length genome alignments at high read depth. Moreover, we show the specificity of the method enables the study of viral population structures and their diversity within a range of clinical samples types.

Citation: Depledge DP, Palser AL, Watson SJ, Lai IY-C, Gray ER, et al. (2011) Specific Capture and Whole-Genome Sequencing of Viruses from Clinical Samples. PLoS ONE 6(11): e27805. doi:10.1371/journal.pone.0027805

Editor: Ravi Jhaveri, Duke University School of Medicine, United States of America

Received June 14, 2011; **Accepted** October 25, 2011; **Published** November 18, 2011

Funding: This work was supported by the Medical Research Centre [G0900950], Wellcome Trust [WT081703MA] and the European Community's Seventh Framework Programme [FP7/2007–2013] under the project EMPERIE, EC grant agreement number 223498. DPD is funded by the Medical Research Council Centre for Molecular Medical Virology [G07008], JB and ERG receive funding from the UCL/UCLH National Institute for Health Research Comprehensive Biomedical Research Centre. PK, ALP, SJW and IY-CL are funded by the Wellcome Trust Sanger Institute. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have read the journal's policy and have the following conflicts. EL is employed by Agilent Technologies, Inc., and Agilent reagents are used in the research presented in this article. This does not alter the authors' adherence to all the PLoS ONE policies on sharing data and materials.

* E-mail: d.depledge@ucl.ac.uk

Introduction

Whole genome sequencing of viral genomes directly from clinical samples is critically important for identifying genetic variants which cause disease, including those that are under positive selection pressure through interaction with the host [1]. Genetic variation defines virus population structures and is used effectively in determining transmission chains [2]. In clinical samples, viral genome copies per millilitre can number in the billions yet the relative proportion of viral nucleic acid is minute in comparison to host nucleic acid. Direct sequencing of mixed human and viral nucleic acids yields representative proportions of sequence reads that map to viral genomes [3], This represents a significant issue when dealing with samples that contain low proportions of viral nucleic acid and one that has limited such studies from being carried out previously [4,5,6,7]. For this reason, current methods for viral genome sequencing benefit significantly from isolation of viral nucleic acid from host nucleic acid prior to sequencing. The primary methods rely on the production of microgram quantities of viral nucleic acid by either *in vitro* virus culture or amplification of virus genomes by PCR [4,5,6,7]. However, both methods are known to alter virus population structures either by replication advantages of subsets of viruses during *in vitro* culture or through the introduction of nucleotide mutations, gene deletions and genome rearrangements [8,9]. Moreover, the presence of PCR-inhibitory secondary structure and the inability of many viral species to thrive in culture present additional difficulties in generating sufficient quantities of viral nucleic acid for whole genome sequencing. These factors all impact on the accuracy of assembled genome sequences and the interpretation of minority population structures.

Some of the hardest viral genomes to sequence are those of the herpesviridae, a family of large highly cell associated (120–230 kilo base pairs (kbp)) DNA viruses. The eight known human-infective Herpesviruses are currently represented by 29 whole genome sequences in GenBank. 18 of these represent Varicella-Zoster Virus (VZV) strains, the causative agent chickenpox and shingles while Epstein-Barr and Kaposi's sarcoma-associated herpes viruses (EBV and KSHV, respectively) are represented by only two strains each. In all cases, these genomes were sequenced using viral nucleic acid isolated from cultured material [7,10,11,12,13, 14,15,16,17,18]).

Target isolation by hybridisation and subsequent enrichment has proven highly effective in exome sequencing studies [19], enabling researchers to target and deep-sequence specific regions within the human genome. This method uses overlapping 120-mer biotinylated RNA baits, designed by tiling across targeted genomic regions. Subsequent hybridisation of the RNA baits with sequence library-prepared nucleic acid enables isolation and enrichment of target material (using a minimal number of rounds of PCR) and generating sufficient quantities for sequencing on second-generation platforms (Illumina, Roche, Abi). Moreover, while microgram quantities of nucleic acid are still required for sequence library preparation, the target genomes need only comprise a fraction of the total nucleic acid [20].

We describe here, the use of a solution-based target capture methodology to separate and enrich for specific viral genomes from low volume clinical samples comprising complex nucleic acid mixtures (including excess human and bacterial nucleic acids). We use a variety of approaches to determine the optimal method for generating sufficient total nucleic acid for sequence library preparation including whole genome amplification methods and the use of carrier nucleic acid. The utility of the method is demonstrated by directly sequencing 13 human herpesvirus genomes from a range of clinical samples including blood, saliva, vesicle fluid, cerebrospinal fluid and tumour cell lines.

Results and Discussion

Initially, total DNA was extracted from a range of VZV, EBV and KSHV clinical and cultured samples (Table 1 and Table S1 online) and their viral loads determined. Due to the decreased sensitivity of the qPCR assay (versus the PCR assay used to confirm presence of viral DNA), no viral load data could be determined for six VZV samples which were below the limit of detection. Five samples underwent whole genome amplification (WGA) using the high fidelity Phi29 DNA polymerase and random primers to generate sufficient DNA for the library preparation step [21]. Viral load assays, post-WGA, showed a large increase in viral nucleic acid within the samples (Table S1). All remaining samples were prepared without WGA, either directly (all culture samples and clinical sample Vesicle I) or with the addition of carrier DNA (clinical samples Blood I). Sequence library preparation, hybridisation and subsequent enrichment were carried out on all samples using the SureSelect Target Enrichment System (Agilent Technologies) [20] and custom designed RNA baits. For comparison, two cultured samples were amplified by overlapping long PCR and the products mixed in equimolar ratios prior to sequence library preparation. The viral load and human DNA content was determined for each sample at the pre-hybridisation, post-hybridisation and post-amplification stages and are expressed as a ratio (Table 1).

All samples were multiplexed (2-7 per lane) and sequenced using a Genome Analyser IIx (Illumina, Inc) yielding between either 4.8×10^7–7.2×10^7 76bp paired-end reads per sample (clinical and cultured samples) or 2.7×10^7–3.3×10^7 54 bp paired-end reads (long PCR amplicons). Post-sequencing, read-pair quality control was performed using QUASR (http://sourceforge.net/projects/quasr/), and removing duplicate and low quality read-pairs. Consensus genome sequences were produced by aligning read-pairs against a reference genome using the Burrows-Wheeler Aligner [22] while polymorphic loci (including SNPs) were reported using VarScan [23]. The accuracy of SNPs identified in the assembled consensus sequences for culture samples I and II and clinical samples Vesicle II and CSF I was confirmed by either direct PCR and sanger sequencing from the original material or prior reporting of the SNP in peer-reviewed publications [24,25] (Table S2). In agreement with previous studies, there was no evidence of error-induced substitutions or indels in the consensus sequences of samples prepared using the Phi29 DNA polymerase for WGA [26].

BLASTn [27] searches of unmapped read-pairs showed them to of human or bacterial origin with minimal homology (<30% identity) to the target enrichment probes, their presence attributed to cross-hybridisation and insufficiently stringent post-hybridisation washes. For samples prepared using the SureSelect system, 34–99% of read-pairs mapped to the reference genomes enabling the generation of full genome consensus sequences (Figure 1 and Table S1). No correlation was observed between viral load and the proportion of mapped reads. Several known short repetitive sequences within the VZV, KSHV and EBV genomes could not be accurately assembled with the BWA algorithm and are not considered further. Genome coverage was lower for samples prepared by long PCR than for target enriched sample. At mapping depths of > 5x per nucleotide, genome coverage was 94–98% for long PCR-prepared samples, compared with > 99% for target enriched samples. At mapping depths of >100x per nucleotide, genome coverage reduced to 88–92% for long PCR samples and ≥ 94% for target enriched samples (Figure S1). These differences are due to the presence of PCR-refractory regions within the VZV genome which have no effect upon the target separation and enrichment method. The specificity of the target enrichment probe sets was confirmed by our ability to specifically target and isolate either KSHV or EBV from a Primary Effusion cell line lysate infected with both viruses using independent RNA bait sets (Table 1). The successful enrichment of viral DNA in each sample is shown by the significant increase in the ratio of viral:human DNA post-hybridisation and is further evidenced by the high proportion of sequence reads that map to the target genome (Table 1).

Minority viral variants have been shown to be important in RNA viruses and there is evidence that diverse population structures among these viruses are strongly associated with viral evolution, disease progression and treatment failure [28,29]. While large DNA viruses are believed to exhibit minimal genetic variation, neither the frequencies of minority variants, nor their biological importance, are known. To examine this in VZV (one of the most stable of the human herpesviruses), we defined polymorphic loci as positions at which a minor allele was present at a frequency between 5–50%, the total read depth exceeded 100 fold and a minimum of 5 independent reads carry the minor allele (Figure 2). By plotting the frequencies of each minority allele, relative to the consensus allele, we generated a 'mutational spectrum' for each sample showing that polymorphic loci exist at between ~0.03–0.5% of positions in the genome (Figure 3). The frequency of VZV genome positions with minority bases was highest in two genomes (Culture III & IV) prepared by long PCR and these also showed strong bias towards A to G and T to C substitutions at minority variant positions, consistent with sequence errors introduced by *Taq*–like polymerases [30]. In contrast, no mutational pattern emerged in any samples prepared by target enrichment confirming that no systemic bias was present. For target enriched samples, those that underwent culture (Culture I and II) had the lowest numbers of minority variant positions (~ 40–50) while the clinical samples were more variable. This likely reflects a generalised tissue culture-related loss of diversity in culture samples [8] while the relatively large proportion of polymorphic loci in CSF I may be indicative of a more diverse population structure, the significance of which is currently unknown.

These data demonstrate, for the first time, the suitability of target capture technology for purifying very low quantities of viral nucleic acid from complex DNA populations where the host genome is in vast excess. This enables deep sequencing and accurate alignment of full length viral genomes directly from clinical samples using next generation technologies, making it far superior to the culture and PCR–based methodologies. The method is sample sparing (compared to traditional techniques), compatible with WGA methods, automatable and applicable to a range of other virus genome types, including RNA viruses. We predict that the method is fully extendable to other pathogens including bacteria and protozoa present in both clinical and environmental samples. Moreover, the ability to recover multiple viral genomes from a single clinical sample using pools of different virus family capture probes offers the potential for next generation multiplex genome sequence based diagnostic testing and studies of host pathogen interactions.

Table 1. Deep sequencing of clinical samples prepared using the SureSelect Target Enrichment System.

	Sample	Starting material	Sample type	Manipulation	Ratio of Viral DNA: Human DNA			% Paired-end reads mapped	%Genome coverage		Mean read Depth per base
					Pre-hybridisation	Post-hybridisation	Post-amplification		>5-fold	>100-fold	
VZV	Culture I	3 µg	Zoster Vaccine Rash	low passage culture	nd	nd	nd	78.66	99.81	98.27	1672
	Culture II	3 µg	Zoster Vaccine Rash	low passage culture	nd	nd	nd	93.98	99.85	98.85	2720
	CSF I	3 µg	Encephalitis	WGA	nd	nd	nd	34.87	99.94	98.28	729
	Vesicle IV	3 µg	Zoster Vaccine Rash	WGA	10299	1157666	9713604	93.69	99.30	97.54	3022
	Saliva I	3 µg	Wild-type Zoster	WGA	2	14	nd	40.15	99.19	94.72	950
	Vesicle III	3 µg	Zoster Vaccine Rash	WGA	34976	1006398	3931100	60.47	99.83	97.88	2416
	Vesicle II	3 µg	Zoster Vaccine Rash	WGA	519875	9855143	856279	96.01	100.00	98.84	1096
	Blood I	250ng*	Wild-type Zoster	none	2	nd	105545	71.14	99.82	97.51	1819
	Vesicle I	500ng	Wild-type Varicella Rash	none	1097	38	nd	99.48	99.93	99.27	3197
EBV	JSC1	2 µg	PEL cell line reactivated virus	culture supernatant	nd	nd	nd	69.10	99.34	98.56	2523
	HBL6	2 µg	PEL cell line reactivated virus	culture supernatant	nd	nd	nd	52.84	98.25	97.17	2599
KSHV	JSC1	4 µg	PEL cell line reactivated virus	culture supernatant	nd	nd	nd	92.01	99.73	95.47	2471
	HBL6	5 µg	PEL cell line reactivated virus	culture supernatant	nd	nd	nd	90.97	98.19	93.92	1773

nd – not determined due to insufficient sample available |
2750ng carrier DNA added.
doi:10.1371/journal.pone.0027805.t001

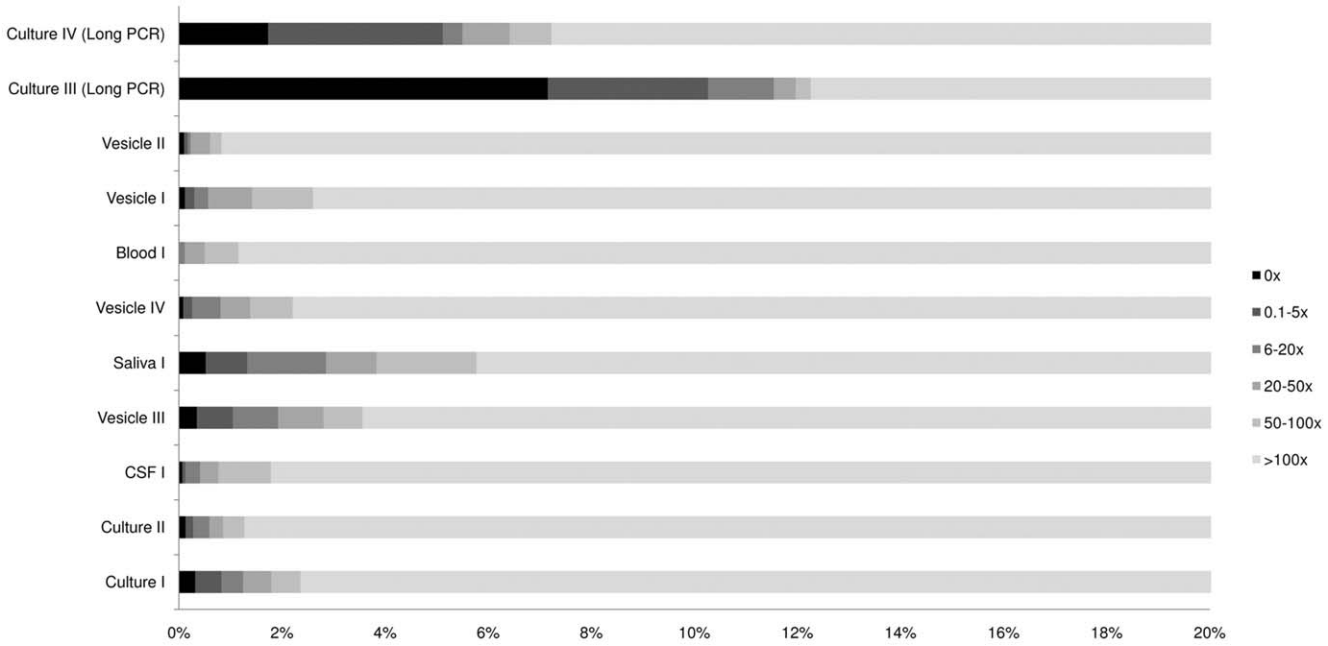


Figure 1. Coverage across sequenced genomes is highest using the target enrichment methods. Proportions of assembled genomes at which read depth per base falls below 100 fold (lightest grey), 50 fold, 20 fold, 5 fold, 1 fold and 0 (indicated by increasing darkness).
doi:10.1371/journal.pone.0027805.g001

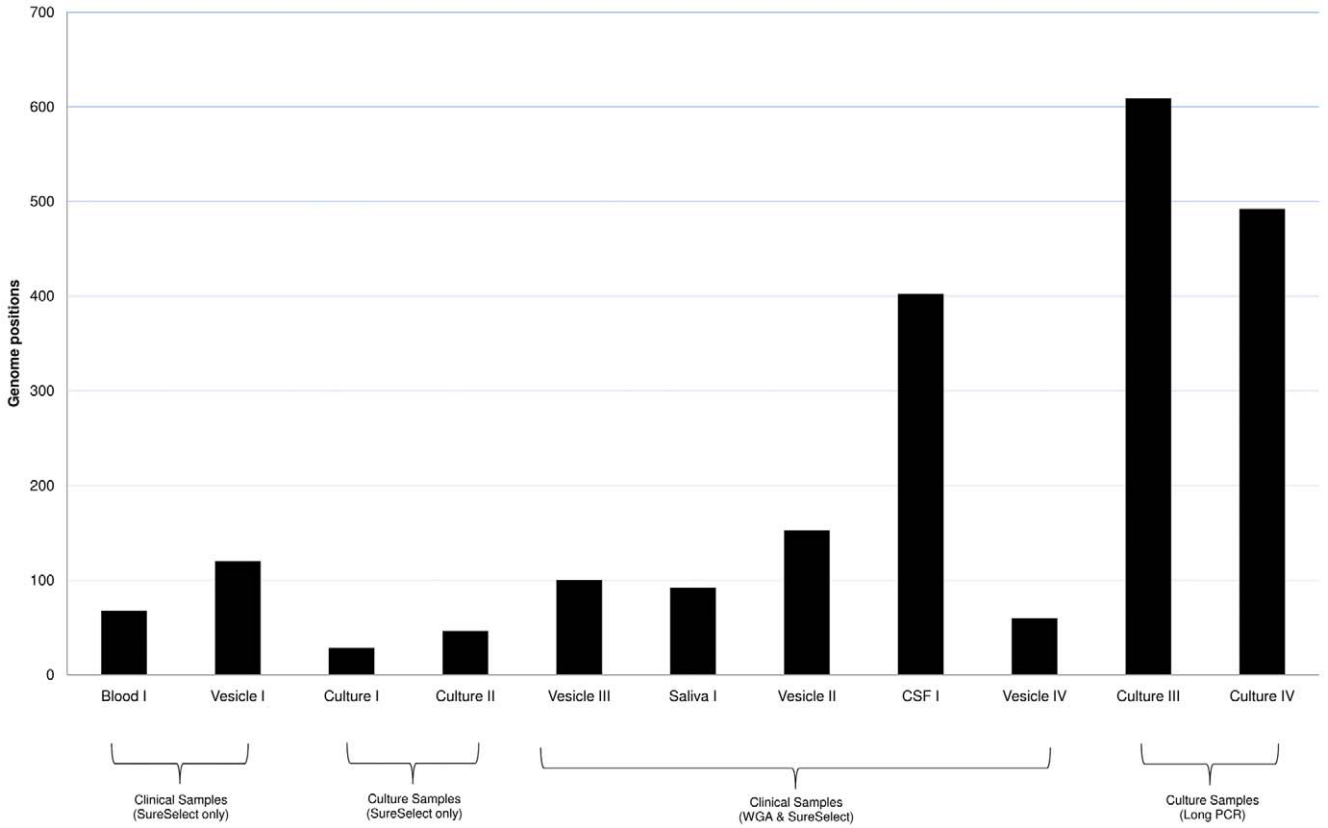


Figure 2. Total numbers of minority variant positions in all sequenced VZV samples. Each bar indicates the number of genome positions at which multiple alleles are present (minor allele frequency 5–49.9%). Datasets are normalised (corrected for the total number of mapped reads per sample) and showed no evidence that minority reads map to specific regions of the genome or that any bias between the proportions occurring in coding and non–coding regions of the genomes is present. Viral genome copies, post-target enrichment could not be determined for some samples (nd).
doi:10.1371/journal.pone.0027805.g002

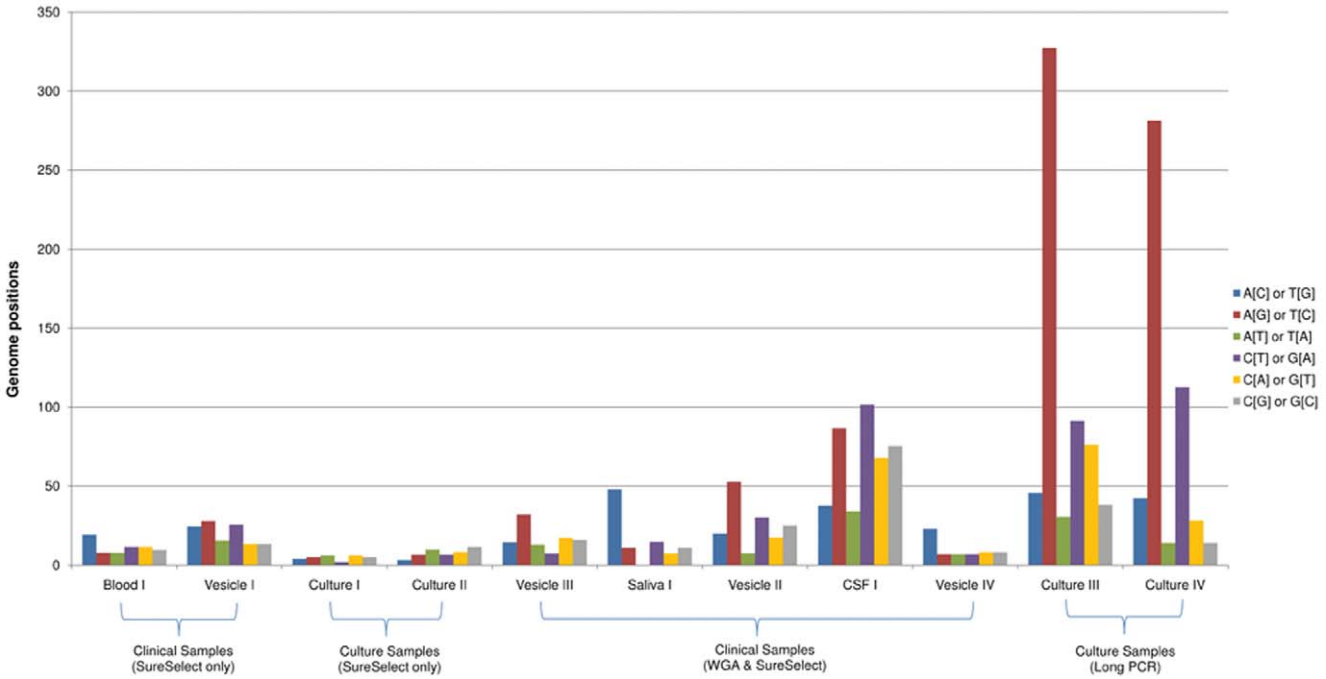


Figure 3. Mutational spectra of minority variants occurring within clinical samples. Each bar indicates the number of genome positions at which specific allele combinations (see graphic) are present (minor allele frequency 1–10%). Datasets are normalised (corrected for the total number of mapped reads per sample) and show a clear bias toward A to G and T to C substitutions in samples prepared by long PCR. No bias was observed in samples prepared using target enrichment methods.
doi:10.1371/journal.pone.0027805.g003

Materials and Methods

Ethics statement

Clinical specimens (diagnostic samples collected as part of standard clinical procedures) were independently obtained from patients with confirmed VZV infection and anonymised prior to this study. Written consent was obtained in all cases. The use of these specimens for research was approved by the East London and City Health Authority Research Ethics Committee (P/96/046: Molecular typing of cases of varicella zoster virus).

Repository of sequence read datasets

All VZV sequence datasets are available in the Sequence Read Archive under the accession number SRA030888.1. All EBV and KSHV datasets are available in the European Genome Archive under the accession EGAS00001000141.

Sample preparation: VZV culture samples

VZV strains Culture I, II, III and IV were retrieved from the Breuer Lab Biobank and cultured (2 passages) in Mewo cells (MEM, 10% FCS, 1% Non-essential amino acids) at 34°C, 5% CO2 until 70–80% cytopathic effect was observed. The monolayer was scraped and centrifuged at 20 0g for 5 min and DNA was extracted using a QiaAmp DNA mini kit (Qiagen) according to manufacturer's instructions.

Sample preparation: VZV diagnostic samples

Diagnostic samples from patients with confirmed VZV infection were retrieved from the Breuer lab cryobank and included vesicle fluid (Vesicle I, II, III and IV), Cerebro-spinal fluid (CSF I) and saliva (Saliva I) and 2 samples adapted to culture (Culture I & II).

Total DNA was isolated from vesicle fluid, saliva and CSF using a QiaAMP DNA mini kit according to manufacturer's instructions. Peripheral blood mononuclear cells (PMBCs) were purified from whole blood samples by centrifugation (1600 g, 15 minutes) enabling separation of plasma (top layer) and PBMCs (middle layer) from red blood cells (bottom layer) and total DNA extracted using a QIAamp DNA Blood Mini Kit according to manufacturer's instructions. Total DNA quantities were determined by NanoDrop and those with a 260/280 ratio outside the range 1.9–2.1 were further purified using the Zymoclean Genomic DNA Clean & Concentrator™ (Zymo Research Corp.).

Sample preparation: Primary effusion lymphoma cell lines

PEL cell lines JSC-1 [31] and HBL6 [32] were cultured in RPMI containing 10% FCS (Biosera) and pen/strep (100 units ml^{-1} penicillin and 100 μg ml^{-1} streptomycin, Invitrogen). Lytic reactivation of KSHV and EBV in PEL was induced by addition of valproic acid (2.5 mg μl^{-1}) and 20 ml virus-containing supernatant collected and 0.45 μm filtered after 72 hours. Viruses were concentrated using 8% Poly(ethylene glycol) triphenylphosphine (Sigma) and 0.15M NaCl. Samples were stored at 4°C for 12 hours before centrifuging (4°C, 2000 g for 10 min). The supernatant was removed and discarded and the virus pellet re-suspended into 200 μl PBS and DNA extracted using the QiaAmp DNA Blood Mini Kit (Qiagen) according to manufacturer's instructions.

Whole genome amplification

5 clinical samples with very low total DNA quantities (with variable viral loads) were amplified (10ng starting DNA) using Genomiphi V2 (GE Healthcare) and purified using Zymoclean Genomic DNA Clean & Concentrator™ (Zymo Research Corp.), both according to manufacturer's instructions.

Viral load assays

The relative proportions of human and viral DNA within each sample were determined by qPCR assays targeted at human GTPase KRas (KRAS) and varicella-zoster virus ORF 29.

VZV DNA was measured by a real-time PCR assay used to quantitatively detect viral DNA in clinical specimens. The PCR targets a 78 bp region in ORF 29 of the VZV genome, a 78 bp region in the EBV nuclear antigen leader protein and a 88 bp region in KSHV ORF 73. For VZV, 1 µl of sample DNA was diluted with 8 µl nuclease-free water and mixed with 12.5 µl of Qiagen master mix (from Quantitect Multiplex PCR Kit (Qiagen)), 0.94 µl (final concentration 0.94 µM) of the forward primer 5′ CACGTATTTTCAGTCCTCTTCAAGTG 3′, 0.94 µl of the reverse primer 5′ TTAGACGTGGAGTTGACATCGTTT 3′ and 0.1 µl of the FAM probe 5′ FAM-TACCGCCCGTGGAGCGCG -BHQ1 3′ (final concentration 0.4 µM). For EBV, the EBNA-LP gene was targeted and samples were prepared with the SensiMix dU kit (Bioline) using a 5 mM MgCl2 concentration, forward and reverse primers at a 20 pmolar final concentration (forward primer 5′ GGCCAGAGGTAAGTGGACTTTAAT 3′, reverse primer 5′ GGGGACCCTGAGACGGG 3′) and a probe at a 10 pmol final concentration (5′ FAM-CCCAACACTCCACCACACCCAGGC-BHQ1 3′). For KSHV, ORF 73 was targeted and samples were prepared as for EBV using the following primers and probe (Forward primer: 5′ TTGCCACCCACGCAGTCT 3′, Reverse primer: 5′ GGACGCATAGGTGTTGAAGAGTCT 3′, Probe: 5′ FAM-TCTTCTCAAAGGCCACCGCTTTCAAGTC-TAMRA 3′) [33]. Quantitative PCR was performed in a 96 well plate on an ABI 7300 or a Masterplex thermocycler ep (Eppendorf) with an initial 15 minute incubation at 95°C followed by 45 cycles at 95°C for 15 seconds and 60°C for 60 seconds. Ct values were compared to a standard curve generated using a plasmid target to assign a copy number per microliter. For human DNA, GTPase KRas was targeted using forward (5′ GCCTGCTGAAAATGACTGAATATAAAC 3′) and reverse (5′ TGATTCTGAATTAGCTGTATCGTCAAG 3′) primers at a 20pmolar final concentration. The relative proportion of human and viral DNA copy numbers was subsequently calculated and expressed as a ratio (Table 1).

SureSelect Target Enrichment: RNA bait design

Overlapping 120-mer RNA baits (generating a 2x coverage for VZV and 5x coverage for EBV and KSHV) spanning the length of the positive strand of the reference genomes were designed using in house Perl scripts for VZV and Agilent eArray software for KSHV and EBV. For VZV, a further 552 control baits were designed against a 16 kbp region of the *Salmo trutta trutta* mitochondrion (NC_010007). The specificity of all baits was verified by BLASTn searches against the Human Genomic + Transcript database. Bait libraries for EBV, KSHV and VZV were uploaded to E-array and synthesised by Agilent Biotechnologies. All bait designs are available from the corresponding author.

SureSelect Target Enrichment: Library preparation, hybridisation and enrichment

DNA preparations of 3 µg, 500 ng and 250 ng (the latter bulked with 2750 ng carrier DNA from MeWo cells) were sheared for 6×60 seconds using a Covaris E210 (duty cycle 10%, intensity 5 and 200 cycles per burst using frequency sweeping). End repair, non-templated addition of 3′-A, adaptor ligation, hybridisation, enrichment PCR and all post- reaction cleanup steps were performed according to the SureSelect Illumina Paired-End Sequencing Library protocol (Version 1.0) observing all recommended quality control steps.

Long PCR

Amplicons ranging from 1–6 kbp in size and spanning the whole VZV genome were generated for culture strains 79A and V110A. 30 overlapping primer pairs were designed against the Dumas reference genome (NC_001348) as a template (Table S3). All reactions were performed using the LongAmp® *Taq* PCR Kit (NEB) and all PCR products size selected by gel purification with the QIAquick Gel Extraction Kit (Qiagen) on 0.8% 1X TAE gels stained with ethidium bromide. Cycling conditions were as follows: Denaturation at 94°C for 3 min, followed by 45 cycles of amplification (denaturation 94°C, 10 s; annealing 55°C, 40 s; extension 65°C, 30 s – 5 m) and a final extension step at 65°C for 10 min. Gel purified amplicons were merged in equimolar ratios prior to library preparation. Sequencing libraries were subsequently generated using the Nextera Tagmentation system (Epicentre Biotechnologies). Here, 50 ng of each sample was sheared and library prepped for paired end sequencing (2×54 bp) in a single reaction according to the manufacturer's instructions. Samples were tagged using the Nextera Barcode Kit and multiplexed prior to flow cell preparation and cluster generation.

Illumina sequencing

Sample multiplexing (2 – 7 samples per lane on an 8 lane flow cell) cluster generation and sequencing was conducted using an Illumina Genome Analyzer IIx (Illumina Inc.) at UCL Genomics (UCL, London, UK) or Wellcome Trust Sanger Institute (Hinxton, UK). Base calling and sample demultiplexing were performed using the standard Illumina pipeline (CASAVA 1.7) producing paired FASTQ files for each sample.

Sequence data processing and alignment against reference genomes

For each data set, all read-pairs were subject to quality control using the QUASR pipeline (http://sourceforge.net/projects/quasr/) to first trim the 3′ end of reads (to ensure the median Phred quality score of the last 15 bases exceeded 30) and subsequently to remove read-pairs if either read had a median Phred quality score below 30 or were less than 50 bp in length. Duplicate read-pairs were also removed. All remaining read-pairs were mapped to the reference genome using the Burrows-Wheeler Aligner (maximum insert 50 bases, maximum distance between paired ends 500) [34] generating SAM files containing all mapped and unmapped reads. SAM files were subsequently processed using SAMTools [35] to produce pileup files for consensus sequence generation and SNP calling using VarScan v2.2.3 (--min-coverage 3, --min-reads2 3, --p-value 5e-02) [23]. Unmapped read-pairs were extracted from SAM files and BLASTn searches used to determine the proportion mapping to the reference genome [27]. Read-pairs with no significant hits were subsequently checked against the non-redundant database at NCBI to determine their origin.

Supporting Information

Figure S1 Mean read depth across assembles genomes. The mean read depth of each position in the assembled genome is shown for (a) VZV culture samples, (b) VZV clinical samples prepared without WGA, (c) VZV clinical samples prepared with WGA, (d) VZV long PCR samples, (e) EBV and KSHV from JSC1 cell lines and (f) EBV and KSHV rom HBL6 cell lines.
(TIF)

Table S1 Deep sequencing of clinical samples prepared using the SureSelect Target Enrichment System.
(DOCX)

Table S2 Confirmation of fixed SNPs identified in assembled consensus sequences.
(DOCX)

Table S3 Primers used to generate overlapping amplicons by long PCR for deep–sequencing of VZV.
(DOCX)

Acknowledgments

The authors would like to thank Prof Paul Kinchington (University of Pittsburgh), Prof Anne Gershon (University of Columbia, New York) and Dr Kate Ward (University College London Hospital Diagnostic Virology lab) for providing all clinical samples used in these studies. We also thank Claire Deback, Meleri Jones and Julianne Lockwood (Infection and Immunity, UCL) for help with sample preparation, Darren Marjenberg (Agilent Technologies) for technical expertise on the SureSelect Target Enrichment System and Justyna Osinska (UCL Genomics) for Illumina sequencing. The authors acknowledge the infrastructure support provided the MRC Centre for Molecular Medical Virology and the use of the UCL *Legion* High Performance Computing Facility, and associated support services, in the completion of this work.

Author Contributions

Conceived and designed the experiments: JB PK DPD RKK ALP. Performed the experiments: DPD ALP IY-CL ERG PG. Analyzed the data: DPD ALP IY-CL SJW JB PK. Contributed reagents/materials/analysis tools: EL ERG DPD. Wrote the paper: DPD JB PK.

References

1. Herbeck JT, Rolland M, Liu Y, McLaughlin S, McNevin J, et al. (2011) Demographic processes affect HIV-1 evolution in primary infection before the onset of selective processes. J Virol.
2. Kew O, Morris-Glasgow V, Landaverde M, Burns C, Shaw J, et al. (2002) Outbreak of poliomyelitis in Hispaniola associated with circulating type 1 vaccine-derived poliovirus. Science 296: 356–359.
3. Liu P, Fang X, Feng Z, Guo YM, Peng RJ, et al. (2011) Direct sequencing and characterization of a clinical isolate of Epstein-Barr virus from nasopharyngeal carcinoma tissue using next-generation sequencing technology. J Virol 85: 11291–11299.
4. Renzette N, Bhattacharjee B, Jensen JD, Gibson L, Kowalik TF (2011) Extensive genome-wide variability of human cytomegalovirus in congenitally infected infants. PLoS Pathog 7: e1001344.
5. Takayama M, Takayama N, Inoue N, Kameoka Y (1996) Application of long PCR method of identification of variations in nucleotide sequences among varicella-zoster virus isolates. J Clin Microbiol 34: 2869–2874.
6. Loparev VN, Gonzalez A, Deleon-Carnes M, Tipples G, Fickenscher H, et al. (2004) Global identification of three major genotypes of varicella-zoster virus: longitudinal clustering and strategies for genotyping. J Virol 78: 8349–8358.
7. Peters GA, Tyler SD, Grose C, Severini A, Gray MJ, et al. (2006) A full-genome phylogenetic analysis of varicella-zoster virus reveals a novel origin of replication-based genotyping scheme and evidence of recombination between major circulating clades. J Virol 80: 9850–9860.
8. Tyler SD, Peters GA, Grose C, Severini A, Gray MJ, et al. (2007) Genomic cartography of varicella-zoster virus: a complete genome-based analysis of strain variability with implications for attenuation and phenotypic differences. Virology 359: 447–458.
9. Dargan DJ, Douglas E, Cunningham C, Jamieson F, Stanton RJ, et al. (2010) Sequential mutations associated with adaptation of human cytomegalovirus to growth in cell culture. J Gen Virol 91: 1535–1546.
10. Davison AJ, Scott JE (1986) The complete DNA sequence of varicella-zoster virus. J Gen Virol 67 (Pt 9): 1759–1816.
11. Dolan A, Addison C, Gatherer D, Davison AJ, McGeoch DJ (2006) The genome of Epstein-Barr virus type 2 strain AG876. Virology 350: 164–170.
12. Glenn M, Rainbow L, Aurade F, Davison A, Schulz TF (1999) Identification of a spliced gene from Kaposi's sarcoma-associated herpesvirus encoding a protein with similarities to latent membrane proteins 1 and 2A of Epstein-Barr virus. J Virol 73: 6953–6963.
13. Gomi Y, Sunamachi H, Mori Y, Nagaike K, Takahashi M, et al. (2002) Comparison of the complete DNA sequences of the Oka varicella vaccine and its parental virus. J Virol 76: 11447–11459.
14. Grose C, Tyler S, Peters G, Hiebert J, Stephens GM, et al. (2004) Complete DNA sequence analyses of the first two varicella-zoster virus glycoprotein E (D150N) mutant viruses found in North America: evolution of genotypes with an accelerated cell spread phenotype. J Virol 78: 6799–6807.
15. Loparev VN, Rubtcova EN, Bostik V, Tzaneva V, Sauerbrei A, et al. (2009) Distribution of varicella-zoster virus (VZV) wild-type genotypes in northern and southern Europe: evidence for high conservation of circulating genotypes. Virology 383: 216–225.
16. Norberg P, Liljeqvist JA, Bergstrom T, Sammons S, Schmid DS, et al. (2006) Complete-genome phylogenetic approach to varicella-zoster virus evolution: genetic divergence and evidence for recombination. J Virol 80: 9569–9576.
17. Tillieux SL, Halsey WS, Thomas ES, Voycik JJ, Sathe GM, et al. (2008) Complete DNA sequences of two oka strain varicella-zoster virus genomes. J Virol 82: 11023–11044.
18. Zeng MS, Li DJ, Liu QL, Song LB, Li MZ, et al. (2005) Genomic sequence analysis of Epstein-Barr virus strain GD1 from a nasopharyngeal carcinoma patient. J Virol 79: 15323–15330.
19. Varela I, Tarpey P, Raine K, Huang D, Ong CK, et al. (2011) Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. Nature 469: 539–542.
20. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, et al. (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol 27: 182–189.
21. Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, et al. (2002) Comprehensive human genome amplification using multiple displacement amplification. Proc Natl Acad Sci U S A 99: 5261–5266.
22. Bonanni P, Breuer J, Gershon A, Gershon M, Hryniewicz W, et al. (2009) Varicella vaccination in Europe - taking the practical approach. BMC Med 7: 26.
23. Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, et al. (2009) VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics 25: 2283–2285.
24. Quinlivan ML, Gershon AA, Al Bassam MM, Steinberg SP, LaRussa P, et al. (2007) Natural selection for rash-forming genotypes of the varicella-zoster vaccine virus detected within immunized human hosts. Proc Natl Acad Sci U S A 104: 208–212.
25. Breuer J, Quinlivan M, Al Bassam M, Macdonald S, Nichols RA, et al. (2007) DNA sequence variability in Oka vaccine isolates. J Infect Dis 196: 801-802; author reply 802-803.
26. Pugh TJ, Delaney AD, Farnoud N, Flibotte S, Griffith M, et al. (2008) Impact of whole genome amplification on analysis of copy number variants. Nucleic Acids Res 36: e80.
27. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, et al. (2009) BLAST+: architecture and applications. BMC Bioinformatics 10: 421.
28. Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R (2006) Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. Nature 439: 344–348.
29. Zagordi O, Klein R, Daumer M, Beerenwinkel N (2010) Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies. Nucleic Acids Res 38: 7400–7409.
30. Bracho MA, Moya A, Barrio E (1998) Contribution of Taq polymerase-induced errors to the estimation of RNA virus diversity. J Gen Virol 79(Pt 12): 2921–2928.
31. Cannon JS, Ciufo D, Hawkins AL, Griffin CA, Borowitz MJ, et al. (2000) A new primary effusion lymphoma-derived cell line yields a highly infectious Kaposi's sarcoma herpesvirus-containing supernatant. J Virol 74: 10187–10193.
32. Gaidano G, Cechova K, Chang Y, Moore PS, Knowles DM, et al. (1996) Establishment of AIDS-related lymphoma cell lines from lymphomatous effusions. Leukemia 10: 1237–1240.
33. Bourboulia D, Aldam D, Lagos D, Allen E, Williams I, et al. (2004) Short- and long-term effects of highly active antiretroviral therapy on Kaposi sarcoma-associated herpesvirus immune responses and viraemia. Aids 18: 485–493.
34. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754–1760.
35. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.